PROPERTIES OF C-LINKED C8-PHENOXYL GUANINE DNA ADDUCTS

ANDREA MILLEN
B.Sc. First Class Honours, Mount Allison University, 2006

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfillment of the
Requirements for the Degree

DOCTOR OF PHILOSOPHY

Department of Chemistry & Biochemistry
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

Dedicated to my father, Clinton, and mother, Linda, my first and best teachers

Abstract

DNA damage is important to understand since it has the potential to lead to disease if unrepaired. In particular, bulky C8 guanine adducts (addition products) are known to induce a variety of mutations due to their conformational flexibility. C-linked C8-phenoxyl-deoxyguanosine adducts (PhOH-dG) have been poorly understood despite their potential for genotoxicity. This thesis systematically develops a computational model to predict the conformational and base-pairing preferences of PhOH-dG by gradually increasing the size of the system. The structure of PhOH-dG in DNA is determined, where the bulky C8 group induces a *syn* conformation of the base similar to other C8-adducts. A stabilized guanine mismatch is identified for the *syn* adducts, which implies that the primary mechanism of genotoxicity may be base-substitution mutations resulting in G→C transversions. This thesis has contributed to a growing body of literature dedicated to understanding the role of conformational heterogeneity in the mutagenicity of bulky C8-adducts.

Acknowledgements

I would like to express my gratitude to my supervisor, Dr. Stacey Wetmore, for providing me with this incredible opportunity, and for years of support and encouragement along the way. I will appreciate it always. I would also like to thank my committee members, Dr. Marc Roussel and Dr. Igor Kovalchuk, for many stimulating discussions in the past five years. To my external examiner, Dr. Chérif Matta, thank you so much for agreeing to participate in this process despite the distance; your opinion is highly respected, and your interest is appreciated. Thanks to the Natural Sciences and Engineering Research Council (NSERC), Alberta Innovates – Technology Futures, and the University of Lethbridge for student scholarships.

I would like to express my deepest thanks to my collaborators, Dr. Richard Manderville and his students (especially Katie Schlitt and Chris McLaughlin), who performed the bulk of the experimental work discussed in this thesis. Special thanks to Richard for his intellectual contributions to this project and many inspiring conversations regarding the direction of this work. Thank you to our system administrators, Jeremiah Merkl and Marc Moreau, for their immense help and support. I would also like to thank the many members of the Wetmore lab, past and present (Aditya Chhikara, Cassandra Churchill, Ken Hunter, Emmanuel Naziga, Jennifer Przybylski, and Lesley Rutledge), for challenging and supporting me throughout my degree. To my undergraduate co-authors from the Wetmore lab, Lex Navarro-Whyte, Cassandra Churchill, Breanne Kamenz, and Fern Leavens, it was a pleasure working with such talented students; thank you. Special thanks to fellow graduate students Cassandra, Jenn, and Lesley for their friendship, and for providing such a stimulating and supportive environment in which to learn. Keeping pace with you has been both a challenge and a pleasure; I know you will go on to do great things.

Finally, accomplishments would not mean much without people to share them with and I have been fortunate to have many. To my friends and family, a special thank you. Your love and support has carried me through the past five years. To my parents, Clinton and Linda, your high standards and deep belief in the value of education inspired me early on. Thank you for encouraging me to achieve this goal. To my sisters, Alisha, Chelsey, and Catherine, much love. To my grandmother, Moi, thanks for the good advice and reminders to take breaks. Thanks to my late grandfather, Baa, for his love and support. To Dot and Ted, thanks for following my progress from the beginning. To Lesley, thanks for listening and being such a good friend, and for many memories I will keep long after I leave here. To Scott, thanks for being there, and getting me through that final stretch. I love you all.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

3,4-EQ, 3,4-estronequinone

4-OHEN, 4-hydroxyequilenin

A, adenine

AA, aromatic amine

AAF, 2-acetylaminofluorene

ABP, aminobiphenyl

AF, 2-aminofluorene

AFB, aflatoxin B_1

AFB-FAPY, aflatoxin B_1 formamidopyrimidine

AN, aniline

AP, aminopyrene

B[a]P, benzo[*a*]pyrene

BPQ, benzo[*a*]pyrene-7,8-dione

C, cytosine

C8-(4-CH_2OCH_3-Ph)-dG, C8-(4-methoxymethylphenyl)-2′-deoxyguanosine adduct

C8-(4-Me-Ph)-dG, C8-(4-methylphenyl)-2′-deoxyguanosine adduct

C8-Ph-dG, C8-phenyl-2′-deoxyguanosine adduct

dA, 2′-deoxyadenosine

dAMP, 2′-deoxyadenosine 5′-monophosphate

dC, 2′-deoxycytidine

dCMP, 2′-deoxycytidine 5′-monophosphate

deg.; degree

DFT, density functional theory

dG, 2′-deoxyguanosine

dGMP, 2′-deoxyguanosine 5′-monophosphate

DMSO, dimethyl sulfoxide

DNA, deoxyribonucleic acid

dT, 2′-deoxythymidine

dTMP, 2′-deoxythymidine 5′-monophosphate

ESIPT, excited state intramolecular proton transfer

G, guanine

HRP, horse radish peroxidise

IARC, International Agency for Research on Cancer

IQ, 2-amino-3-methylimidazo[4,5-*f*]quinoline

MD, molecular dynamics

N2-(4-PhOH)-dG, N2-(4-hydroxyphenyl)-2′-deoxyguanosine adduct

NI, 5-guanidino-4-nitroimidazole

NMR, nuclear magnetic resonance

o-, *ortho*

ODN, oligodeoxynucleotide

OTA, ochratoxin A

p-, *para*

PAH, polycyclic aromatic hydrocarbons

PCM, polarisable continuum model

PCP, pentachlorophenol

PDB, Protein Data Bank

Ph, phenyl

PhIP, 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine

PhOH, phenoxyl

RESP, restrained electrostatic potential

T, thymine

ZPVE, zero-point vibrational energy

1 Chapter 1: Introduction

1.1 Thesis Overview

Deoxyribonucleic acid (DNA) contains the genetic information of a cell, and its complementary structure ensures that genetic information is passed down accurately from generation to generation. However, the integrity of the genetic code is constantly threatened by cellular processes that alter or degrade the chemical structure of DNA. Since it is of the utmost importance that genetic information remains intact, organisms have evolved intricate DNA repair processes that help ensure alterations to the DNA structure are identified and removed before replication. In the event that these repair processes fail, mutations and permanent alteration of the genetic code can result, which can initiate apoptosis or the carcinogenic process. This thesis develops a computational model to understand the properties of a form of DNA damage induced by phenoxyl radicals. The following sections detail the structure of DNA, mutagenic mechanisms and associated biological consequences for a variety of DNA lesions, and a summary of the contents of this thesis.

1.2 DNA Structure

DNA is composed of two anti-parallel phosphodiester backbones interwound in a helical fashion and held together by hydrogen-bonding interactions via four possible nucleobases. The nucleobases (Figure 1.1) are divided into two groups: the monocyclic pyrimidines (cytosine (C) and thymine (T), Figure 1.1a) and the bicyclic purines (adenine (A) and guanine (G), Figure 1.1b). Attachment of C1′ in the deoxyribose sugar to the nucleobases via a C–N (glycosidic) bond (N1 for the pyrimidines and N9 for the purines) composes the deoxynucleosides (deoxycytidine (dC), deoxythymidine (dT), deoxyadenosine (dA) and deoxyguanosine (dG), Figure 1.2a). The deoxyribose sugar can be extended to include a 5′ or 3′ phosphate group, which yields a deoxynucleoside 5′-monophosphate

(Figure 1.2b) or deoxynucleoside 3′-monophosphate (Figure 1.2c). These three fundamental parts (base, sugar, phosphate) compose a nucleotide, which is the repeating unit of a DNA strand.

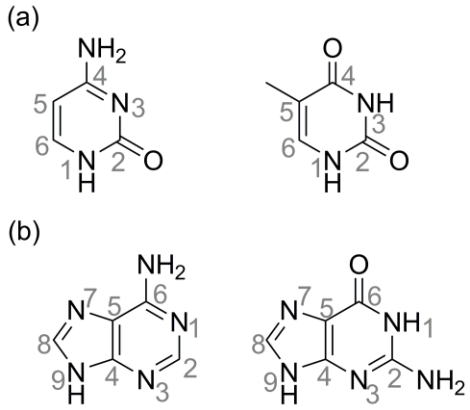


Figure 1.1 Structure and numbering of the four natural DNA nucleobases. (a) The pyrimidines, cytosine (left) and thymine (right) and (b) the purines, adenine (left) and guanine (right).

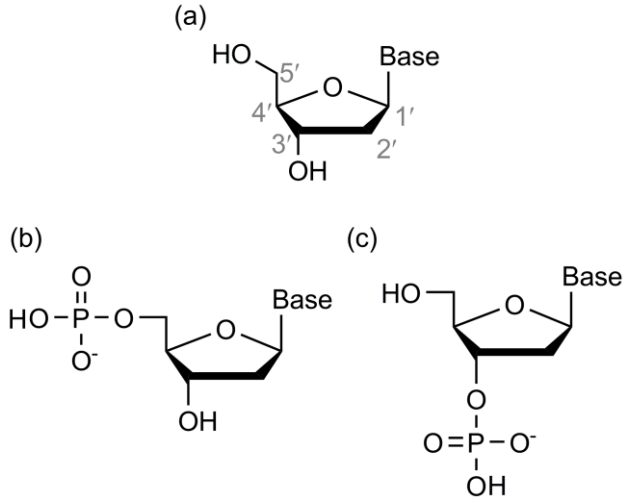


Figure 1.2 Structure and numbering of the deoxyribose sugar which together with the base compose a (a) nucleoside, (b) nucleoside 5′-monophosphate and (c) 3′-monophosphate.

The purines and pyrimidines form specific complementary base pairs via Watson-Crick hydrogen-bonding interactions between A and T (Figure 1.3a), or G and C (Figure

1.3b). This requires the nucleobases to adopt the *anti* conformation about the glycosidic bond (Figure 1.4, left), which is defined by the dihedral angle $\chi = 180 \pm 90°$ ($\angle$(O4′C1′N9C4) for purines and $\angle$(O4′C1′N1C2) for pyrimidines). When the base rotates about the glycosidic bond to adopt the *syn* conformation ($\chi = 0 \pm 90°$ Figure 1.4, right), the Hoogsteen bonding face becomes available for base pairing. In DNA, adoption of the *syn* conformation rather than *anti* can lead to mutations due to the loss of Watson-Crick base pairing and the possibility of Hoogsteen mispairing, and is usually the result of a damage event.

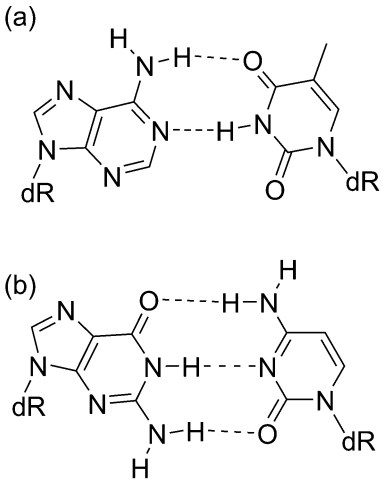


Figure 1.3 The Watson-Crick hydrogen-bonded base pairs between (a) adenine and thymine, and (b) guanine and cytosine.

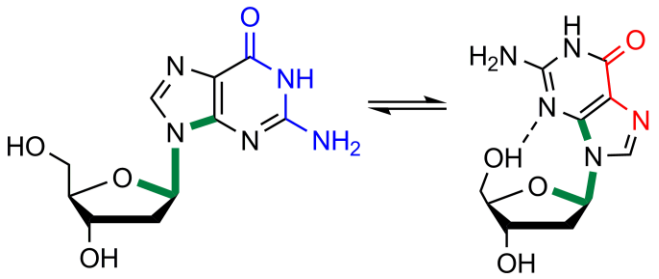


Figure 1.4 The *anti* conformation of the glycosidic bond (left, $\chi = 180 \pm 90°$) with the Watson-Crick face (blue) and the *syn* conformation (right, $\chi = 0 \pm 90°$) with the Hoogsteen face (red) exposed. The dihedral angle χ dictates the conformation of the base with respect to the sugar (green (bold)).

The most common DNA helix conformation is B-DNA, which is characterized as a right-handed helix with a 36° average twist and 10 base pairs per turn.[1,2] A-DNA (right-handed helix with 11.6 base pairs per turn) is of minor importance in a biological context.[2] A third rare helix variety, Z-DNA, exists, but conversion of B-DNA to Z-DNA involves a drastic conformational change, which only occurs in helices that contain alternating purine and pyrimidine sequences. Specifically, the helix transforms to a left-handed twist with 12 base pairs per turn, where the purines adopt the *syn* conformation and the pyrimidines remain in the *anti* conformation.[2] Z-DNA is hypothesized to play a role in gene expression.[2]

1.3 DNA Mutations

During replication, the DNA helix unwinds and each strand becomes a template to be read by a polymerase enzyme. The polymerase synthesizes a daughter strand by extending a primer sequence paired with the template strand. Specifically, an incoming nucleotide is inserted opposite each base in the template, and then phosphodiester bonds are formed between the incoming nucleotide and the previous nucleotide in the daughter strand. The polymerase uses sterics and hydrogen-bonding interactions to determine which of the four bases will best complement the template base. Normally, DNA polymerases show remarkable accuracy and selectivity for the correct base, with an error rate as low as 1 per10^7 bases (depending on the type of polymerase).[3] However, even subtle changes to the structure of a base can result in loss of recognition by the polymerase and subsequent misincorporation of the incoming nucleotide.[1]

When polymerases fail, two broad classes of mutations are possible. The first is base-substitution mutations,[3] which result from changes to the physical features of DNA bases such as heteroatom placement, presence of tautomers and induced conformational changes that distort the normal duplex structure.[1] Specifically, heteroatom placement and the presence of tautomers can alter normal Watson-Crick hydrogen bonding, which may

result in nucleobase transversions (substitution of a purine for a pyrimidine, or vice versa) or transitions (substitution of a purine for a purine, or pyrimidine for a pyrimdine) during the replication process.[2] Changes to the base can also affect the stability of the glycosidic bond,[4,5] resulting in bond cleavage and formation of an abasic site.[4,5] This leads to base-substitution mutations when the polymerase pairs the abasic site with an incorrect base (often adenine).[3]

The second class of mutations is frameshift mutations, where an altered base (and possibly any neighbors affected by the damage) is altogether skipped by a polymerase.[6] This results in deletions, where a full base pair (or more) is lost upon replication of the daughter strand. Insertions are another form of frameshift mutation, where an extra base (or more) is inserted into the daughter strand. Frameshift mutations are believed to result from chemical changes to DNA bases that stabilize the formation of bulges, which are DNA defects consisting of one or more unpaired bases in either complementary strand,[6] and can form during replication as intermediate mutagenic structures.[7]

Base-substitution and frameshift mutations are associated with diseases such as cancer.[8,9] Therefore, it is important to understand the DNA damaging events that lead to these disruptions in DNA replication. In particular, this thesis will investigate the structural effects of phenoxyl radical-induced damage at the C8 site of dG. The following section will discuss a variety of DNA damaging events and the biological implications, before focusing on the phenoxyl damage of interest in this thesis.

1.4 DNA Damage

DNA is under constant assault due to exposure to endogenous and exogenous species which can cause mutations if unrepaired. Nucleobase oxidation is a common occurrence in cells, where for example, hydroxyl radicals that are byproducts of metabolism can attack the C8 site of guanine and lead to the premutagenic 8-oxoguanine lesion (Figure 1.5a), which

miscodes with adenine upon replication due to favorable Hoogsteen bonding. Although the other nucleobases are also affected by oxidation damage, 8-oxoguanine is the most common product due to the low oxidation potential of guanine and the vulnerability of the C8 site to radical attack.[10] Ultimately, 8-oxoguanine leads to G→T transversion mutations if unrepaired prior to replication. Nucleobase deamination is another common damaging event in cells. In particular, cytosine is highly susceptible to hydrolytic deamination at the N4 site, which forms uracil (Figure 1.5b). The uracil lesion has a similar hydrogen-bonding pattern to thymine, and is therefore highly miscoding for adenine during replication. If unrepaired, uracil leads to C→T transition mutations.[11] Alkylation damage is also common, where, for example, the environmental carcinogen vinyl chloride can damage adenine to produce 1,N6-ethenoadenine (Figure 1.5c).[12] This compound blocks polymerases during replication, and primarily leads to miscoding of A if incorporation is successful.[12] Thus, A→T transversion mutations can occur. It is important to identify and understand mutagenic lesions, such as base oxidation, deamination, and alkylation, since the accumulation of mutations can lead to tumor growth and cancer.[10,13]

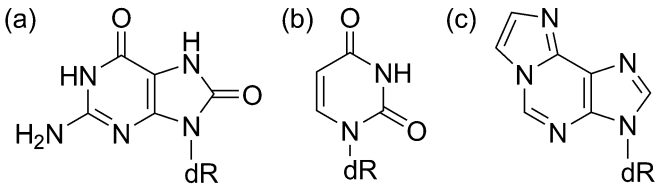


Figure 1.5 Examples of mutagenic DNA lesions: (a) 8-oxoguanine, (b) uracil, and (c) 1,N6-ethenoadenine.

This thesis focuses on a type of DNA nucleobase damage known as a bulky addition product, or "adduct," where a bulky aryl group covalently bonds to a nucleobase. The following section describes the importance of understanding DNA adducts and details a variety of sources of this type of damage. An emphasis is placed on damage to the guanine

nucleobase, particularly at the C8 site (C8-dG), since the structures of adducts covalently bonded to this site are most relevant to the phenoxyl damage studied in this thesis.

1.4.1 Bulky DNA Adducts

Bulky DNA adducts occur when chemicals composed of one or more aromatic rings are consumed and metabolized to reactive intermediates that interact with DNA. Many environmental contaminants have been identified as possible genotoxic agents that result in bulky DNA adducts. However, it can be difficult to determine the exact mechanism of toxicity of these chemicals and prove their genotoxic activity. Without sufficient evidence that an environmental contaminant is a genotoxin, such as proof of a defined DNA adduct structure, legislation that limits the extent of exposure to humans via, for example, dietary intake, water, and personal products, is significantly relaxed.[14] Therefore, the study of the exact structure of DNA damage products and the possible mutagenic mechanism of compounds considered to be of potential risk is important from the standpoint of influencing legislative attitudes,[14] as well as from the perspective of developing preventive and therapeutic measures for combating the carcinogenicity of such compounds.

The mutational spectra of DNA adducts are explained by a variety of theories. Historically, it was hypothesized that each individual adduct results in a specific mutation.[15] However, there is increasing evidence that single adducts lead to a variety of mutations depending on the sequence context.[16] Thus, it was proposed that multiple mutations are induced when a single adduct adopts a variety of conformations, which is further complicated by the number of ways each conformation can interact with a polymerase during replication.[15,16] Since C8-dG adducts have been observed to result in mixtures of mutations, it has been proposed that each mutation could be conformation specific.[16] In the case of bulky C8-dG adducts, multiple conformations arise due to the increased steric bulk at the C8 position, which may result in nucleobase rotation about the glycosidic bond to

yield a more energetically favourable *syn* conformation rather than the native *anti* conformation,[17] as in the case of 8-oxoguanine. Thus, the conformational flexibility of C8-dG adducts can cause pronounced perturbation of the helix and lead to mixtures of base-substitution and frameshift mutations if unrepaired. The factors that control the conformational equilibrium are therefore important to study for these adducts. It has been found that the preferred conformation is dependent on a number of factors, including adduct type, site of attachment (C8, N2, N7 or O6 of G), identity of flanking bases and complementary base, and environment (replication fork, polymerase active site, etc.). Therefore, it is important to study each conformation in order to understand possible associated mutations.

Environmental chemical substances are classified as human carcinogens by the World Health Organization's International Agency for Research on Cancer (IARC).[9] The compounds described in the following sections are ubiquitous environmental chemicals that are either known or possible human carcinogens due at least in part to their genotoxic activity. The following section will describe the DNA adducts that have been identified for each species, with an emphasis on C8-dG adducts, and the role conformational changes induced by the adduct play in their genotoxicity. The section begins with a discussion of what is currently known about adducts formed by phenoxyl radicals, which are the focus of this thesis.

1.4.1.1 Phenoxyl Adducts

Some substituted phenols, such as pentachlorophenol, are classified as environmental carcinogens by the IARC, in part due to their ability to cause oxidative stress and DNA damage in cells,[18] including DNA adduct formation.[18-21] Other chlorophenol compounds that have been identified as potential carcinogens are common environmental contaminants found in pesticides, disinfectants, wood preservatives, and are by-products of

wood pulp bleaching with chlorine.[18] Ochratoxin A (OTA) is a *para*-chlorophenolic mycotoxin produced by several species of *Aspergillus* and *Penicillium* fungi.[14] Although the growth of OTA-producing molds is not readily controlled in stored animal and human foods,[14] the toxic effects of OTA exposure include carcinogenicity, mutagenicity, and nephrotoxicity,[22] and therefore this compound is classified as a possible human carcinogen by the IARC.[18]

Studies show OTA leads to carbon-bonded C8-dG adducts (Figure 1.6) following oxidative activation.[18] For pentachlorophenol, the predominant pathway for metabolic activation leads to formation of the *o*-quinone species, which forms adducts at the N2/N1 site of dG.[18] However, recent studies show that pentachlorophenol can also form a phenoxyl radical intermediate that results in stable, oxygen-bonded C8-dG adducts (Figure 1.6).[21] Thus, substituted phenoxyl compounds can display an ambident reactivity at the C8 site of dG, where chlorophenol radicals tend to form oxygen-bonded dG adducts[21] and more heavily-substituted phenols, such as ochratoxin A, form *para*-C-bonded C8-dG adducts.[23]

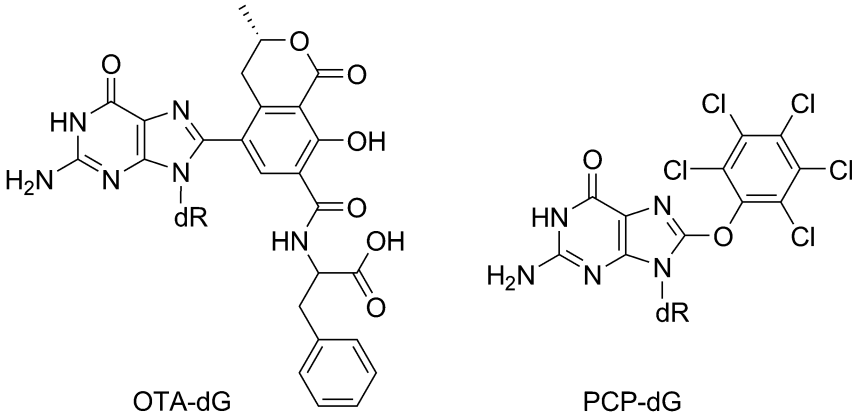


Figure 1.6 Structure of the *para* carbon-bonded ochratoxin A C8-dG adduct (OTA-dG) and the oxygen-bonded pentachlorophenol C8-dG adduct (PCP-dG).

Phenoxyl radical formation in humans can be mediated by peroxidase enzymes or redox-active transition metals.[24,25] Studies show unsubstituted C8-dG phenoxyl adducts

(PhOH-dG) are formed when phenol reacts with excess nitrite to generate diazoquinones that degrade to phenoxyl radicals, which covalently attach to C8 of dG (Figure 1.7).[26,27] Diazoquinones are mutagenic in the Ames test without metabolic activation and it is expected that carbon-bonded C8-PhOH-dG adducts are responsible for the mutagenic activity.[26-29] Phenoxyl adducts have also been implicated in the genotoxic effects of benzene,[30] which can be metabolized to *p*-benzoquinone and hydroquinone.[31] These species predominantly result in formation of *p*-PhOH-dG at the N2 site of dG, rather than C8-adducts.[31]

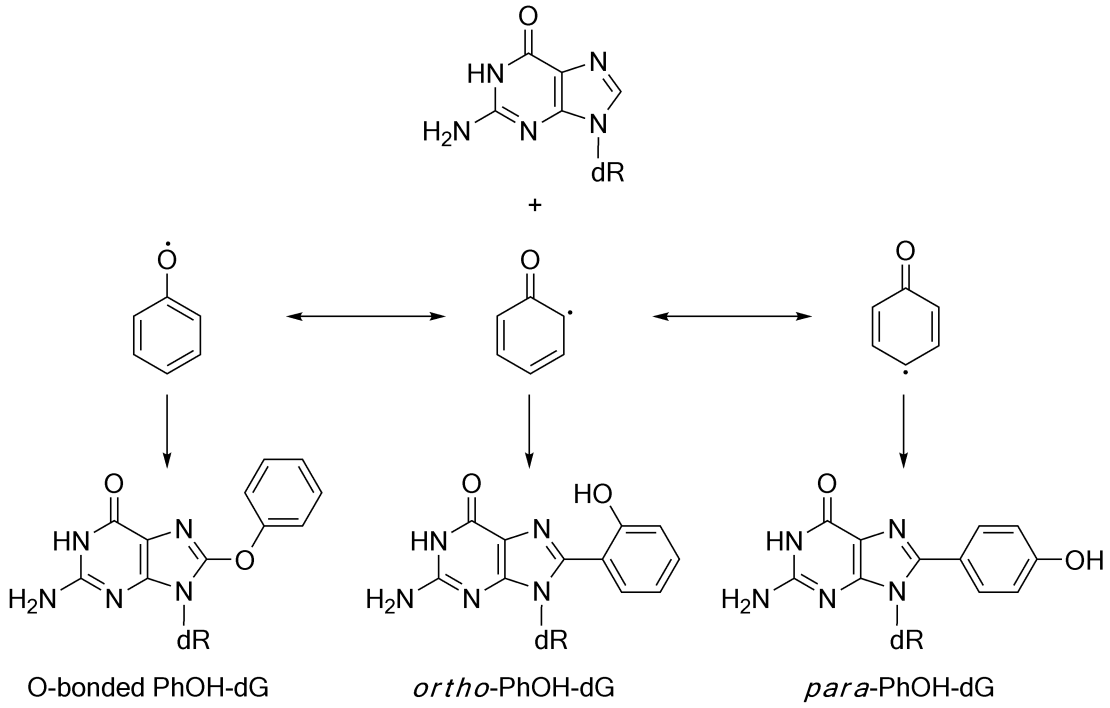


Figure 1.7 Formation of the oxygen- and carbon-bonded C8-PhOH-dG adducts.

The *ortho* (*o*-PhOH-dG) and *para* (*p*-PhOH-dG) carbon-bonded C8-dG phenoxyl adducts (Figure 1.7) have recently been studied experimentally.[18,32,33] Specifically, the structure of *o*-PhOH-dG involves an intramolecular O–H···N7 hydrogen bond, which allows for excited-state intramolecular proton transfer (ESIPT) to occur to generate the keto tautomer when the adduct adopts a planar orientation.[32] Therefore, UV-vis

spectrophotometry can be used as a tool to investigate the ground-state properties of the molecule[18,32,33] by providing clues regarding the adduct conformation based on the presence or absence of the intramolecular hydrogen bond.[32] Specifically, the planar/twisted conformations of the *o*-PhOH-dG adduct were found to be sensitive to the solvent environment,[32] as discussed in Chapter 2. The *p*-PhOH-dG adduct is also of significant experimental interest due to its ability to act as a fluorophore,[33] and therefore, its potential use as a pH-sensing fluorescent probe.[33] Despite the potential importance of both adducts as bioprobes for conformational changes in DNA,[34] many questions remain regarding their structure and their observed solvent-dependent conformational flexibility.[32]

Computational chemistry can help interpret and complement recent experimental findings regarding the solvent-dependent conformational flexibility of the *o*-PhOH-dG adduct and thus allows for a greater overall understanding of phenoxyl damage. This thesis therefore uses calculations to understand the conformational properties and base-pairing preferences of carbon-bonded PhOH-dG adducts, which are comparatively less well-studied than related carcinogenic C8-dG adducts. Although calculations have proven to reveal important information about other bulky adducts,[35-45] no computational work has been done on C8-PhOH-dG adducts. In addition, the structural relationship between the simplest phenoxyl adducts and other more complex C8 bulky adducts, such as those stemming from exposure to aryl hydrazines and estrogens (discussed below), suggests that a thorough study of phenoxyl adducts will help develop the current general understanding of the carcinogenic model of carbon-bonded C8-aryl-dG damage. The following section discusses the literature regarding other bulky adducts to identify the important questions for phenoxyl adducts that remain to be addressed.

1.4.1.2 Aryl Hydrazine Adducts

Humans are exposed to aryl hydrazines through industrial, pharmaceutical, and natural (dietary) sources.[42,46] For example, several aryl hydrazines are found in the mushroom *Agaricus bisporus*, which is sold commonly in grocery stores.[42,46] Therefore, the associated biological effects have been the subject of much interest.[46] It has been determined that the reactive metabolic intermediates formed by these chemicals form reactive oxygen species that damage cells.[46] However, it has also been found that aryl hydrazines with various *para* substituents can be metabolized to arenediazonium ions and then to aryl radicals that form C8-aryl-dG adducts.[26,47-51]

C8-aryl hydrazine adducts (Figure 1.8) closely resemble the structure of *p*-PhOh-dG adducts (Figure 1.7), differing only in the identity of the *para* substituent. The formation of abasic sites has been observed for these adducts, which is proposed to result from a decrease in stability of the glycosidic bond upon adduct formation.[42,48,52] As discussed in Section 1.3, abasic sites result in G→T transversion mutations, and are therefore an important consequence of adduct formation.

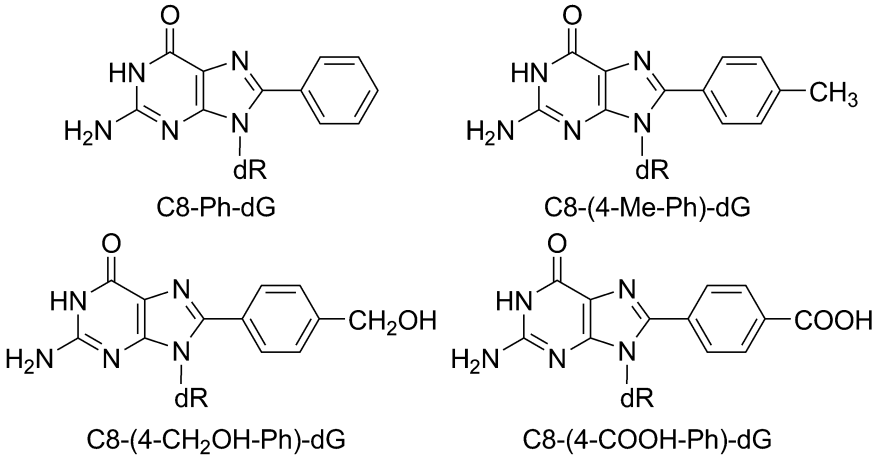


Figure 1.8 Structures of adducts identified from exposure to aryl hydrazines.

Several studies have also investigated the tendency of alternating GC sequences containing aryl hydrazine adducts to undergo a conformational change from B to Z-form DNA.[42,44,53] Since Z-DNA has been implicated in the activation of a gene promoter (CSF1), Z-DNA stabilization may help explain the carcinogenicity of these adducts.[54] Molecular dynamics (MD) simulations performed on C8-Ph-dG, C8-(4-Me-Ph)-dG, and C8-(4-CH_2OCH_3-Ph)-dG indicate an energetic preference for Z- vs. B-DNA in alternating GC sequences, where the adduct adopts a *syn* conformation.[42] However, B- vs. Z-DNA stability does not fully account for the carcinogenicity of the aryl hydrazine compounds, which implies that other factors, such as repair or mutation, play a role in the biological activity.[42] While Z-DNA stabilization has been studied,[44,53] it is important to note that the formation of C8-(4-X-Ph)-dG adducts has also been connected to mutations,[47] where exposure to aryl hydrazines causes mutations in bacteria and tumors in mice.[46,47]

The possibility of base-substitution, as well as frameshift, mutations has been proposed for aryl hydrazine adducts.[47] Mutagenicity of these lesions was tested by site-specifically incorporating C8-Ph-dG into oligonucleotides and using the modified strands as templates in primer extension reactions catalyzed by various DNA polymerases.[47] It was demonstrated that C8-Ph-dG is capable of blocking primer extension. In addition, small amounts of dA and dG misincorporation were observed, as well as one- and two-base deletions.[47] Only low levels of mutations were observed for the C8-Ph-dG lesion, which generated G→T and G→C transversions. However, miscoding was found to be polymerase-dependent.[47] The *syn*/*anti* conformational preference of C8-Ph-dG was speculated to play a role in the biological activity of the adduct,[47] where multiple conformations may be responsible for the variety of mutagenic outcomes.

Although much has been done to study the impact of aryl hydrazine adducts on Z-DNA stability, computational work has yet to investigate the experimental observations of

the mutagenic properties of these adducts. The variety of pathways described for aryl hydrazine adducts to induce mutations and the uncertainty regarding their mechanism of action suggests a need for further study to determine the role these adducts may play in carcinogenesis. The remarkable structural similarity of these adducts to PhOH-dG suggests they may have similar mechanisms of mutagenicity. Therefore, a greater understanding of PhOH-dG adducts will also contribute to knowledge of aryl hydrazine mutagenicity.

1.4.1.3 Aromatic Amine Adducts

Aromatic amines and heterocyclic aromatic amines (AAs) are environmental and dietary contaminants that are well-known DNA damage-inducing agents,[38,40,55-67] since workers in the textile industry who were exposed to various AAs experienced high rates of bladder cancer.[9] Despite their known activity, many aromatic amines can still be found in a variety of sources (color additives, paints, food colors, leather and textile dyes, fumes from cooking oils, fuels, tobacco smoke, hair dye, and smoke from charred meat).[9] Humans are therefore exposed to heterocyclic aromatic amines frequently, and this is associated with a variety of tumor sites.[9] Particularly potent aromatic amines have been selected for study as model carcinogens. Specifically, 2-aminofluorene (AF) and N2-acetylaminofluorene (AAF) (Figure 1.9) were originally developed as pesticides, but were banned before use, due to their carcinogenic nature. The carcinogenicity of these compounds is correlated to the formation of DNA adducts (Figure 1.9).[9] The principal products that occur after exposure to these compounds are nitrogen-bonded adducts at C8 or N2 of guanine, or N6 of adenine.[9]

Three conformational motifs have been identified for the variety of aromatic amine adducts and their derivatives. First, *anti* structures are formed where the bulky group is located in the major groove. Second, *syn* base-displaced, intercalated structures can be generated, where the base opposite the adduct adopts an extrahelical position and the bulky group is stacked with flanking base pairs in the helix.[9,16] Third, another *syn* conformation,

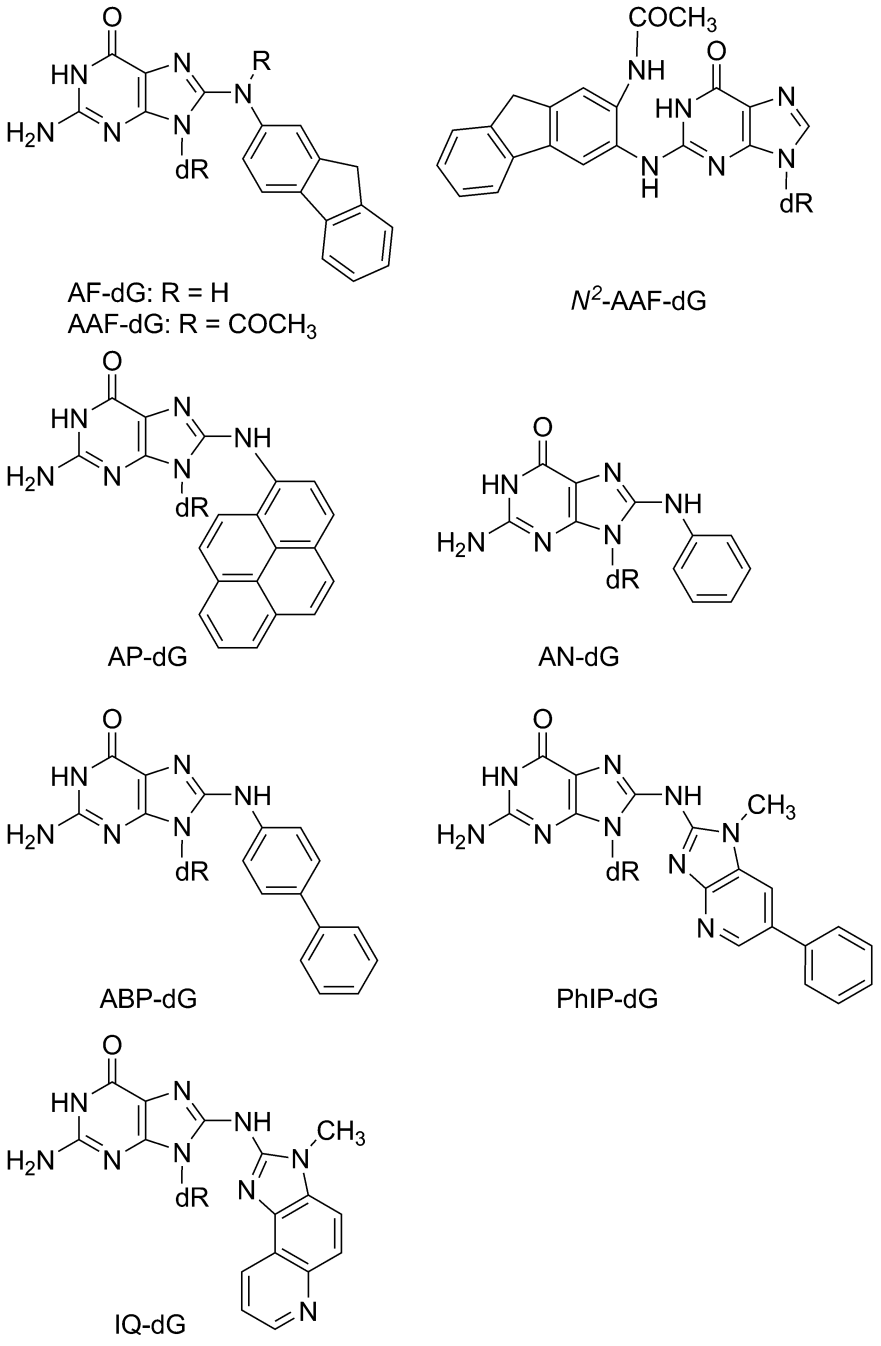


Figure 1.9 Structures of aromatic amine adducts.

known as the wedged conformer, is possible, where the bulky group protrudes into the minor groove and the base opposite is not displaced.[16] Stacking between the bulky group and the flanking bases stabilizes the intercalated structures, whereas Watson-Crick base

pairing is maintained to stabilize the *anti* conformer. It is hypothesized that nondistorting *anti* conformers may escape repair, while distorting *syn* conformers may be mutagenic,[58,68] and therefore the equilibrium between these conformations helps determine the potential mutagenicity. Therefore, an in-depth understanding of the conformations and the forces controlling their relative populations is important for understanding the mutagenicity of bulky adducts.

The physical features of the N-linked adducts help determine which of the three conformations is preferentially adopted. For example, it has been found that adducts with larger bulky groups, such as in AP-dG (Figure 1.9), result in a greater population of the *syn* stacked conformer due to increased stacking stabilization when the bulky group intercalates into the helix and displaces the opposite base. Some studies show that the AAF-dG adduct (Figure 1.9) solely adopts the stacked intercalated conformer.[69] NMR studies reveal that damaged duplexes with deletions opposite AAF-dG (post-replication model) result in greater thermal stabilities than duplexes containing deletions without bulky modifications due to the strength of stacking interactions with the intercalated bulky group.[70,71] This results in bulge stabilization, which explains why AAF-dG leads primarily to frameshift mutations (deletions) upon replication.[72-74] For comparison, the single-ring aniline adduct (AN-dG, Figure 1.9) exclusively adopts the *anti* conformation and distorts B-DNA structure much less than the bulkier AAF-dG adduct.[75]

The flexibility of the bulky group relative to the dG moiety of the adduct also impacts the relative populations of the conformations, where greater coplanarity of the adduct base results in higher levels of the stacked conformation. For example, although AAF-dG and AF-dG have similar sized bulky groups, AF-dG is capable of adopting both the *syn* stacked and *anti* conformers. The lack of conformational exchange for AAF-dG compared to AF-dG is likely due to the presence of the acetyl group at the N-linker site (Figure 1.9), which limits

flexibility due to steric constraints.[9] Furthermore, ABP-dG (Figure 1.9) has a similar structure to AF-dG (lacks one methylene bridge), but has different *anti*/*syn* equilibria since the two phenyl rings can rotate with respect to one another and therefore greater deviations from coplanarity occur than in AF-dG.[76] This results in a 90 – 95% population of the *anti* conformer for ABP-dG[76,77] compared to 55 – 70% for AF-dG. PhIP-dG (Figure 1.9) is also flexible in this manner, with a conformer equilibrium that is temperature-dependent.[9] In contrast, the large co-planar AP-dG adduct (Figure 1.9) almost exclusively forms the stacked conformer.[78,79] The *anti*/*syn* equilibrium of the IQ-dG adduct (Figure 1.9) favors the *syn* conformation.[80,81] However, it can adopt a stacked or wedged *syn* conformer depending on the degree of twist in the N-linker bonds at the site of attachment to C8.[80] It has been suggested that the wedged conformer is preferred over the stacked conformer for IQ-dG because the methyl group disrupts intercalation into the helix.[82]

Other factors, such as sequence context, have also been observed to affect adduct conformation.[9,83] For example, the AF-dG adduct has been reported to have an *anti*/*syn* equilibrium that is dependent on the length and sequence context of the oligomer. Specifically, a 15mer with a TG*A (G* = adduct) sequence contained 55% of the *anti* conformer,[84] while a 10mer with the AG*G sequence contained 60 – 70% of the *anti* conformer.[85,86] The *Nar*I sequence, CTCG^{1}G^{2}CG^{3}CCATC, which is a genetic hotspot for mutations, was also studied with the adduct in each of the G (G^1, G^2, G^3) positions. AF-dG exists in a 50% conformational mixture at the G^3 site, but favors the *anti* conformation by 70 – 90% at G^1 and G^2.[87]

The role of interstrand hydrogen bonding in adduct structure has also been examined. For example, recent studies of primer extension past AF-dG have revealed that pairing with A promotes a persistent *syn* conformation, while pairing with C leads to a transition from *syn* to *anti*.[88] Similar results have been found for other guanine lesions. For

example, MD results[89] have been used to show that the *anti* or *syn* conformation of a guanine lesion can be adopted depending on the partner base. In addition, 8-oxoguanine lesions share a similar conformational flexibility, where the base adopts a *syn* conformation when paired with A, and an *anti* conformation when paired with C in DNA.[90]

The genotoxicity of AF and AAF has been studied extensively via experimental[35,36,40,57,58,64,67,70-74,83,87,88,91-93] and computational methods[35-40] and used as a model for understanding the biological activity of bulky C8 adducts. AF adducts primarily lead to base-substitution mutations.[92] NMR studies have revealed a *syn* conformation where the adduct is Hoogsteen hydrogen-bonded to A in double–stranded DNA.[35] This suggests a possible mechanism for formation of the G→T mutation most frequently observed for AF-dG.[67,94] Alternatively, AAF-dG primarily causes two-base deletions.[72-74] AP-dG (Figure 1.9) adducts also lead to frameshift mutations with two base deletions,[95] as well as transition mutations depending on the cell type and sequence context.[62] IQ-dG predominantly leads to frameshift mutations (base deletions) and to a lesser extent G→C to T→A transversions.[82]

Studies of the aromatic amine adducts have provided much insight into a possible model for understanding C8 adduct-induced mutations. However, much remains to be learned. Additionally, the role of the N-linker bonds in adduct conformation makes it difficult to apply what is known about the AA adducts to C-linked adducts, such as PhOH-dG. There is also a complicated interplay between the size of the bulky group, adduct planarity, sequence, and interstrand hydrogen bonding in determining adduct conformation and ultimately the duplex structure. Therefore, it is important to systematically study each individual adduct conformation and its potential mutations.

1.4.1.4 Polycyclic Aromatic Hydrocarbon Adducts

Humans are exposed to polycyclic aromatic hydrocarbons (PAHs), which are characterized by the presence of two or more fused benzene rings,[9] through environmental

sources, such as fossil fuel combustion.[9] PAHs are a component of fine particulate matter, and are therefore an extremely common contaminant found in the air, soil, water supplies, and the food chain.[9] PAHs are linked to human lung cancer due to their presence in tobacco smoke.[9] For example, the very common PAH compound, B[*a*]P (DNA damage products shown in Figure 1.10), has been classified by the IARC as a carcinogenic substance to humans.[9]

PAHs that act as carcinogens share some common structural features. They generally have four or more fused benzene rings and include a bay region (see, for example, Figure 1.10a).[9,96] When the bay region is either methylated or closed by a fourth ring, a sterically-hindered fjord region is created, which induces nonplanarity in the aromatic ring system.[97] Decreases in planarity are associated with increases in potency of the carcinogens.[97] PAHs can be metabolically activated in humans to form reactive intermediates that directly lead to DNA adducts, which are well-studied in the literature.[16,43,96,98-102]

Known DNA adducts formed from PAH exposure include those generated via a radical cation intermediate in the metabolic pathway,[103,104] such as B[*a*]P-6-C8-dG, B[*a*]P-6-N7-dG, and B[*a*]P-6-N7-dA (Figure 1.10a).[100] These adducts are all proposed to lead to depurination and consequently to mutations associated with abasic site formation,[100,105,106] similar to the C8-aryl hydrazine adducts discussed previously.

In addition to the radical cation pathway, PAHs can also be metabolically activated to form diol epoxides.[100] These species can lead to 16 possible DNA adducts (many of which are stereoisomers), which covalently attach to either N2 of dG or N6 of dA[68] (Figure 1.10b). The N2-dG PAH adducts adopt an *anti* conformation, which places the bulky group in the minor groove similar to the *syn* wedged AA adducts,[98] but with Watson-Crick pairing maintained.[97] However, it is also possible for the bulky group of the *anti* conformer to

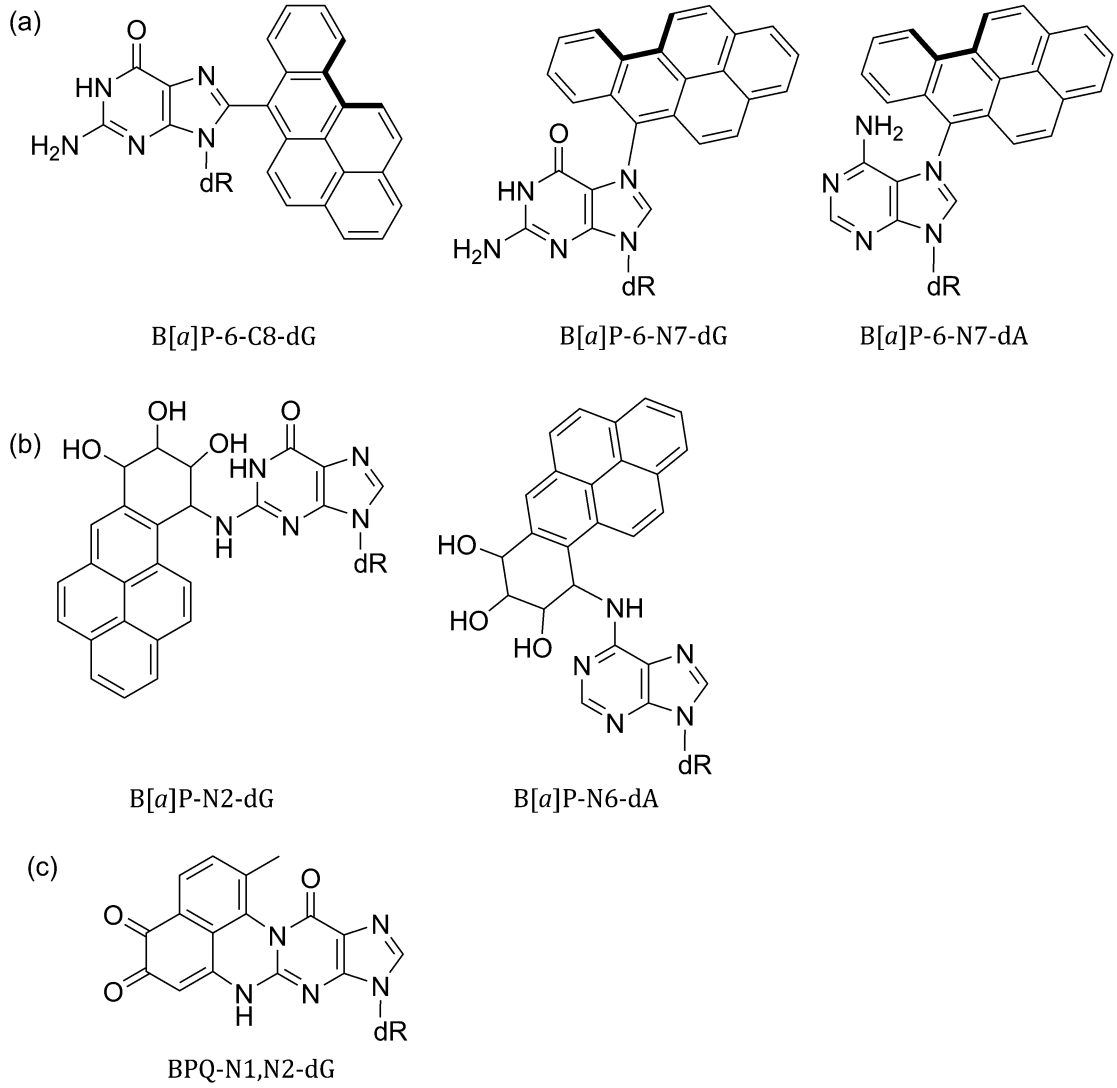


Figure 1.10 Structures of adducts resulting from exposure to the polycyclic aromatic hydrocarbon benzo[*a*]pyrene (stereoisomers not shown). Example of a bay region is highlighted in bold.

intercalate into the helix and displace the opposite base.[97] Therefore, despite the difference in connectivity between the C8-AA-dG and N2-PAH-dG adducts, the bulky group adopts similar conformations in the helix.

PAHs can also be activated by a third pathway to form quinones such as benzo[*a*]pyrene-7,8-dione (BPQ).[9,43] These react to form stable bulky adducts due to the 1,4- or 1,6-Michael addition to the N2 site of dG or the N6 site of dA (Figure 1.10c).[9] The

multitude of DNA adducts due to exposure to even a single PAH compound (Figure 1.10) makes it difficult to study the connection to mutagenicity experimentally. This is further complicated by the complex mixtures of a variety of PAHs that are normally present upon exposure to these chemicals.[9]

In light of the difficulty associated with studying these compounds experimentally, MD studies of PAH adducts have provided insight into their mutagenicity.[41,45] In particular, studies have focused on the conformations adopted by the adducts within the active site of the polymerase paired opposite each of the four bases. This reveals how the steric fit of the incoming nucleotide is affected by the conformation adopted by the lesion.[41,45] Due to the differences in connectivity of PAH compared to AA adducts, the *anti* conformation is generally preferred by all stable N2 PAH adducts.[97] However, the orientation of the bulky group can alternate between stacked (intercalated into the helix with displacement of the opposite base) and wedged (projecting into the minor groove).[97] Like AA adducts, the dominant conformation is dependent on the size, topology, and stereochemistry of the base,[102] as well as the sequence context of the lesion.[102]

Certain PAH-dG adducts are structurally related to PhOH-dG adducts since they are carbon-bonded, can occur at the C8 site (Figure 1.10a), and often contain phenol rings (Figure 1.10b). Therefore, it is extremely important to apply what is known regarding PAH adducts to the study of PhOH adducts, and consider similar mechanisms of mutation for both adduct types. Since several of the depurinating PAH adducts are bonded to the C8 site of guanine, this suggests that the stability of C8-PhOH-dG adducts in DNA should also be addressed in this thesis. In addition, it is interesting to consider the possibility of a variety of mutations for PhOH-dG adducts due to a conformational heterogeneity.

1.4.1.5 Estrogen Adducts

Human estrogens have been classified as carcinogens by the IARC.[9] The primary mechanism of toxicity involves hormonal activity.[9] However, a secondary mechanism has been proposed that involves activation of human estrogens to *o*-quinone intermediates that form DNA adducts.[9] For example, equine estrogen *o*-quinone metabolites, such as equilin and equilenin, which are used in hormone replacement therapy in humans, have been shown to form DNA adducts.[9,107] In particular, the estrogen 3,4-estronequinone (3,4-EQ) undergoes metabolic activation to a 3,4-EQ radical anion, which generates C8-dG adducts such as 8-3,4-EQ-dG[26,108,109] (Figure 1.11). Like C8-bonded PAH and aryl hydrazine adducts, it has been proposed that 8-3,4-EQ-dG is an intermediate to abasic site formation.[108-110] Other estrogens have also been found to predominantly lead to abasic sites after formation of unstable adducts, but stable adducts are a minor product as well, such as 4-OHEN-dG (Figure1.11)[107,110] Relatively little is currently known about these adducts compared to AA and PAH adducts, and their connection to cancer (for example, breast cancer, prostate cancer, and non-Hodgkin lymphoma) is currently under study.[110,111] The structural similarity between C8-3,4-EQ-dG and the PhOH-dG adducts suggests that their mutagenic mechanisms may also be related, and therefore study of phenoxyl adducts in this thesis will be beneficial for understanding estrogen genotoxicity.

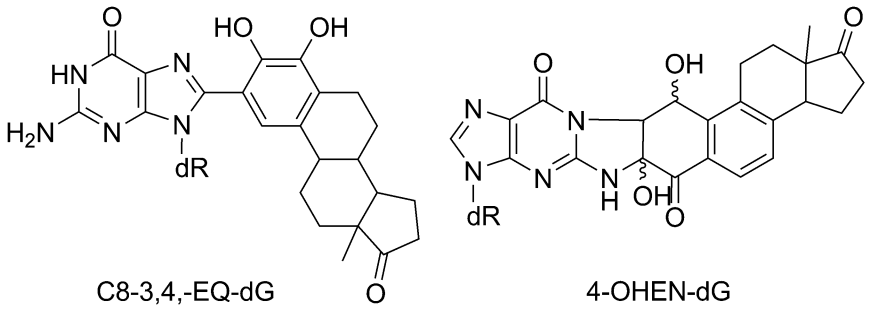
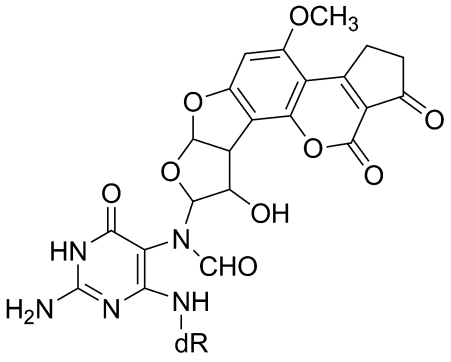


Figure 1.11 Examples of structures of known estrogen adducts.

1.4.1.6 Aflatoxin Adducts

Aflatoxins are extremely carcinogenic mycotoxins.[112,113] They are produced by the *Aspergillus* fungal species that contaminate food crops such as stored grains, and are consumed in the diet.[9] Exposure to aflatoxins is associated with a high rate of human hepatocellular carcinoma.[9] Aflatoxin B_1 (AFB) is metabolized to a highly reactive intermediate that primarily forms N7-dG adducts.[112] While most N7-dG adducts directly lead to depurination,[5] the AFB-dG adduct undergoes ring-opening to form the stable AFB formamidopyrimidine adduct AFB-FAPY-dG[112] (Figure 1.12). This adduct is conformationally complex and the flexible formamide group interconverts between a conformer that causes mutations and one that blocks replication.[112] The preferred conformation is also sequence dependent, where the flexible group can form hydrogen-bonding interactions with neighboring bases.[112] Research into this adduct is ongoing and will contribute to the understanding of how a complex conformational heterogeneity may result in complex mutation spectra, and how discrete interactions with neighboring bases can provide insight into sequence effects on adduct conformation. Therefore, a better understanding of the flexibility of bulky adducts in DNA will provide insight into mutagenic mechanisms of both AFB-FAPY-dG and PhOH-dG adducts.



AFB-FAPY-dG

Figure 1.12 Structure of the N7-dG adduct formed from exposure to aflatoxin B_1.

1.5 Summary

Although relatively little is known about phenoxyl damage, adducts that are closely related in structure to PhOH-dG, such as the aryl hydrazine, AA, PAH, and estrogen adducts discussed above, are known carcinogens, where the conformation adopted by the adduct in DNA is often linked to the mutagenicity. Therefore, it is of interest to address the structural changes incurred by phenoxyl radical attack at C8-dG in detail. In addition, the experimental interest in phenoxyl adducts outlined in Section 1.4.1.1 will benefit from a thorough computational analysis of the structural and biological impact of phenoxyl-mediated DNA damage. Due to the similarity in adduct structure, understanding phenoxyl adducts will also help predict the conformational flexibility of other, substituted C8-phenyl adducts. Thus, this thesis will enhance the development of the carcinogenic model for C8-phenyl adducts in general.

It is interesting that several conformational themes are common among the well-studied C8 aromatic amine and N2 PAH adducts, as this suggests a possibility for similar mechanisms of toxicity for other adducts. While the aromatic amine adducts can adopt either the *anti* or *syn* conformation, PAHs tend to adopt the *anti* conformer in duplex DNA. However, a difference in connectivity for the C8-AA-dG versus the N2-PAH-dG adducts leads to similar conformations of the bulky group in the helix. Specifically, base-displaced, intercalated structures, as well as wedged conformations are observed for both types of adducts. Furthermore, the mutagenicity of the adducts is determined at least in part by the *anti* or *syn* conformation adopted, and the orientation of the bulky group. For the phenoxyl adducts, it is unknown whether the *anti* or *syn* conformation is more accessible, and whether wedged or stacked conformations of the bulky group will be preferred. Therefore, it is important to study the structure and *anti*/*syn* preferences of the phenoxyl adducts. Additionally, it should be noted that the conformations of the phenoxyl adducts may result

in different helical structures than AA or PAH adducts due to the relatively small size of the phenoxyl group. Since conformation and biological activity are related for other bulky adducts, understanding the possible conformations and the factors controlling the relative populations will allow predictions to be made regarding the mutagenicity of the phenoxyl adducts in different environments.

Several physical features of adducts have been identified as important for understanding and predicting the (*anti*/*syn*) conformational preferences of bulky adducts. First, planarity of the bulky group has been correlated to the conformational preference, as well as the mutagenic potency of the adducts. For example, planar bulky groups at the C8 site of dG can cause adducts to preferentially adopt a *syn* stacked conformation where the bulky group intercalates into the helix. In addition, the degree of twist between the dG moiety and the bulky group, which is dictated by the linker bonds to the C8 or N2 site, affects the conformational preference. In the case of phenoxyl adducts, the bulky group is planar, but the phenoxyl ring may be twisted relative to the nucleobase, where the C–C carbon bond connecting the phenyl group to the C8 site controls the degree of twist (Figure 1.8). Therefore, the planarity of phenoxyl adducts is also of interest to study.

The size of the ring system has also been correlated to toxicity, where more rings often means more potent effects. This may imply less toxicity for smaller systems such as the PhOH-dG adducts. For example, while very bulky AA and PAH adducts are capable of stabilizing bulges and thereby inducing frameshift mutations, it is unknown whether the comparatively small size of the phenoxyl adducts will result in similar bulge stabilization. Furthermore, if intercalation of the phenoxyl ring in the helix occurs, an extrahelical conformation of the opposite base as observed for AA and PAH adducts may not result, since the small ring size may readily accommodate the opposite base in the helix. However, it has been suggested that lesions that result in smaller perturbations to helix structure are

less likely to be recognized by repair enzymes,[75] and this may lead to greater mutagenic potential.[68]

It is also interesting that the C8 aryl hydrazine, PAH, and estrogen adducts have been reported to induce mutations via abasic site formation, given their structural similarity to the C8-PhOH-dG adducts. For example, treatment of DNA with B[*a*]P (Figure 1.10) in aqueous media with horseradish peroxidase (HRP)/HOOH afforded mixtures of B[*a*]P-6-C8-dG and the corresponding nucleobase analogue.[114] Significant levels of abasic site formation have also been reported for arylhydrazine treatment of DNA.[52] It has also been proposed that 8-3,4-EQ-dG (Figure 1.11) is formed as an intermediate prior to the loss of the deoxyribose sugar to generate 8-3,4-EQ-8-G.[109] Overall, these results have led to proposals that formation of abasic sites following C8-aryl adduct formation may contribute to the carcinogenicity of PAHs[103,104,106] estrogens,[109] and arylhydrazines.[52] Therefore, abasic site formation should be considered for PhOH-dG adducts, where previous experimental work has studied the acid-catalyzed hydrolysis of a variety of *o*- and *p*-substituted Ph-dG adducts.[115] Calculations can shed light on the relative reactivities of these systems by considering the effects of the bulky C8 group on the stability of the glycosidic bond.

As shown above, structural knowledge of DNA adducts can lead to an understanding of their functional role in the onset of carcinogenesis.[97] Therefore, obtaining oligonucleotide structures that contain a single specific lesion has been the focus of much research with the goal of providing insights into the mechanisms of mutagenesis.[8,16] Computational chemistry is particularly well-suited to this task, where a single conformation can be isolated to study the corresponding mutations, whereas experiments must account for a complex mixture of conformations. Indeed, calculations have proven useful tools for the study of a variety of DNA damage types.[35-45,89] Specifically, these studies provided valuable insight into the structure and conformations of DNA adducts, typically in environments (i.e.; specific

sequences or active sites) for which experimental data is available for comparison. Development of a computational model that incorporates the surrounding environment in a systematic manner will help clarify the important factors dictating the structural preferences of PhOH-dG adducts, for which only limited experimental data is available. This will also allow for comparison between small and large model results to determine the success of small models in predicting adduct conformation and base-pairing preferences in the absence of experimental data. Thus, the most accurate models will be identified for future studies of other C8 adducts when little evidence of the preferred structure is available.

1.6 Thesis Approach and Summary

The present work will develop a computational model for the study of the structure and properties of carbon-bonded C8-dG phenoxyl adducts with the goals of understanding: 1) the conformational flexibility of phenoxyl adducts in a DNA environment, and 2) how phenoxyl adducts may contribute to genotoxicity in cells. With this goal in mind, Chapter 2 will provide a thorough density functional theory (DFT) study of the structure and preferred conformation of the adduct using nucleobase and nucleoside models, with comparison to available experimental data. The possibility that abasic site formation contributes to the mutagenicity of C8-dG adducts will also be addressed through a study of the stability of the glycosidic bond in a variety of *o*- and *p*-substituted C8-Ph-dG adducts. In particular, the structural changes in the adduct upon deglycosylation will be used to explain experimental findings regarding the acid-catalyzed hydrolysis of the adducts. Chapter 3 will use the nucleobase and nucleoside models to predict the base-pairing preferences of the adducts in an effort to determine the potential for mispairing upon replication. Chapter 4 will further analyse the structures of the adducts using nucleoside and nucleotide models to understand the effect of backbone interactions on the conformational preferences. In

Chapter 5, the computational model will be further expanded to a deoxydinucleoside monophosphate to understand the effects of intrastrand flanking bases on the *anti*/*syn* conformational preferences. Finally, in Chapter 6, the results from the small model studies will be used to design a DNA duplex model, where the adduct structure is investigated in two different oligonucleotides. This will provide insight into the effects of the duplex environment on the preferred adduct conformation, where the base-pairing preferences in each environment will also be considered to predict the potential mutations associated with the adducts.

The systematic approach developed to model the structure of phenoxyl adducts allows for comparison between small models and duplexes. Thus, conclusions can be drawn regarding the success of small computational models as predictors for adduct conformation and, consequently, biological implications. Furthermore, the smaller computational models often correspond to systems for which experimental data is available, and can improve interpretation of experimental results. The Conclusions chapter will summarize the key findings of this body of work and recommend computational models for the study of other modified C8-dG bases. In addition, the Future Work section outlines what remains to be done to determine the mutagenicity of PhOH-dG adducts, and how the methodology in this thesis can be used to answer important questions about related systems.

1.7 References

(1) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag New York Inc.: New York, NY, 1984.

(2) Voet, D.; Voet, J. G. *Biochemistry*; 3 ed.; John Wiley & Sons, 2004.

(3) McCulloch, S. D.; Kunkel, T. A. *Cell Res.* **2008**, *18*, 148-161.

(4) Boysen, G.; Pachkowski, B. F.; Nakamura, J.; Swenberg, J. A. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* **2009**, *678*, 76-94.

(5) Gates, K. S.; Nooner, T.; Dutta, S. *Chem. Res. Toxicol.* **2004**, *17*, 839-856.

(6) Minetti, C.; Remeta, D. P.; Dickstein, R.; Breslauer, K. J. *Nucleic Acids Res.* **2010**, *38*, 97-116.

(7) Roy, D.; Hingerty, B. E.; Shapiro, R.; Broyde, S. *Chem. Res. Toxicol.* **1998**, *11*, 1301-1311.

(8) Delaney, J. C.; Essigmann, J. M. *Chem. Res. Toxicol.* **2008**, *21*, 232-252.

(9) Geacintov, N. E.; Broyde, S.; Editors *The Chemical Biology Of DNA Damage*; Wiley-VCH Verlag GmbH & Co. KGaA, 2010.

(10) Nakabeppu, Y.; Tsuchimoto, D.; Yamaguchi, H.; Sakumi, K. *J. Neurosci. Res.* **2007**, *85*, 919-934.

(11) Kreutzer, D. A.; Essigmann, J. M. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 3578-3582.

(12) Hang, B.; Chenna, A.; Guliaev, A. B.; Singer, B. *Mutat. Res.-Fundam. Mol. Mech. Mutagen.* **2003**, *531*, 191-203.

(13) *Biomarkers of Environmentally Associated Disease Technologies, Concepts and Perspectives*; 1st ed.; Wilson, S. H.; Suk, W. A., Eds.; Lewis Publishers, CRC Press LLC: Florida, 2002.

(14) Mantle, P. G.; Faucet-Marquis, V.; Manderville, R. A.; Squillaci, B.; Pfohl-Leszkowicz, A. *Chem. Res. Toxicol.* **2010**, *23*, 89-98.

(15) See, K. Y.; Jelinsky, S. A.; Loechler, E. L. *Mutat. Res.-Rev. Mutat. Res.* **2000**, *463*, 215-246.

(16) Cho, B. S. P. *J. Environ. Sci. Health Pt. C-Environ. Carcinog. Ecotoxicol. Rev.* **2004**, *22*, 57-90.

(17) Eason, R. G.; Burkhardt, D. M.; Phillips, S. J.; Smith, D. P.; David, S. S. *Nucleic Acids Res.* **1996**, *24*, 890-897.

(18) Manderville, R. A. *Can. J. Chem.-Rev. Can. Chim.* **2005**, *83*, 1261-1267.

(19) Dai, J.; Lantero, D. R.; Sloat, A. L.; Adams, M.; Akman, S. A.; Manderville, R. A. *Chem. Res. Toxicol.* **2003**, *16*, 59.

(20) Dai, J.; Sloat, A. L.; Wright, M. W.; Manderville, R. A. *Chem. Res. Toxicol.* **2005**, *18*, 771-779.

(21) Dai, J.; Wright, M. W.; Manderville, R. A. *Chem. Res. Toxicol.* **2003**, *16*, 817-821.

(22) Pfohl-Leszkowicz, A.; Manderville, R. A. *Molecular Nutrition & Food Research* **2007**, *51*, 61-99.

(23) Dai, J.; Wright, M. W.; Manderville, R. A. *J. Am. Chem. Soc.* **2003**, *125*, 3716-3717.

(24) Gaikwad, N. W.; Bodell, W. J. *Chem.-Biol. Interact.* **2003**, *145*, 149-158.

(25) Lin, P. H.; Nakamura, J.; Yamaguchi, S.; La, D. K.; Upton, P. B.; Swenberg, J. A. *Carcinogenesis* **2001**, *22*, 635-639.

(26) Manderville, R. A. In *Advances in Physical Organic Chemistry, Vol 43*; Academic Press Ltd-Elsevier Science Ltd: London, 2009; Vol. 43, p 177-218.

(27) Kikugawa, K.; Kato, T.; Kojima, K. *Mutation Research* **1992**, *268*, 65-75.

(28) Kikugawa, K.; Kato, T. *Food Chem. Toxicol.* **1988**, *26*, 209-214.

(29) Hiramoto, K.; Kojima, K.; Kato, T.; Kikugawa, K. *Mutation Research* **1992**, *272*, 259-260.

(30) Selassie, C. D.; DeSoyza, T. V.; Rosario, M.; Gao, H.; Hansch, C. *Chem.-Biol. Interact.* **1998**, *113*, 175-190.

(31) Rodriguez, B.; Yang, Y. N.; Guliaev, A. B.; Chenna, A.; Hang, B. *Toxicol. Lett.* **2010**, *193*, 26-32.

(32) McLaughlin, C. K.; Lantero, D. R.; Manderville, R. A. *J. Phys. Chem. A* **2006**, *110*, 6224-6230.

(33) Sun, K. M.; McLaughlin, C. K.; Lantero, D. R.; Manderville, R. A. *J. Am. Chem. Soc.* **2007**, *129*, 1894-1895.

(34) Weishar, J. L.; McLaughlin, C. K.; Baker, M.; Gabryelski, W.; Manderville, R. A. *Org. Lett.* **2008**, *10*, 1839-1842.

(35) Norman, D.; Abuaf, P.; Hingerty, B. E.; Live, D.; Grunberger, D.; Broyde, S.; Patel, D. J. *Biochemistry* **1989**, *28*, 7462-7476.

(36) Ohandley, S. F.; Sanford, D. G.; Xu, R.; Lester, C. C.; Hingerty, B. E.; Broyde, S.; Krugh, T. R. *Biochemistry* **1993**, *32*, 2481-2497.

(37) Shapiro, R.; Sidawi, D.; Miao, Y. S.; Hingerty, B. E.; Schmidt, K. E.; Moskowitz, J.; Broyde, S. *Chem. Res. Toxicol.* **1994**, *7*, 239-253.

(38) Wang, L. H.; Broyde, S. *Nucleic Acids Res.* **2006**, *34*, 785-795.

(39) Meneni, S. R.; Shell, S. M.; Gao, L.; Jurecka, P.; Lee, W.; Sponer, J.; Zou, Y.; Chiarelli, M. P.; Cho, B. P. *Biochemistry* **2007**, *46*, 11263-11278.

(40) Donny-Clark, K.; Shapiro, R.; Broyde, S. *Biochemistry* **2009**, *48*, 7-18.

(41) Perlow, R. A.; Broyde, S. *J. Mol. Biol.* **2002**, *322*, 291-309.

(42) Heavner, S.; Gannett, P. M. *J. Biomol. Struct. Dyn.* **2005**, *23*, 203-219.

(43) Jin, B.; Lee, H. M.; Kim, S. K. *J. Biomol. Struct. Dyn.* **2010**, *27*, 457-464.

(44) Vongsutilers, V.; Phillips, D. J.; Train, B. C.; McKelvey, G. R.; Thomsen, N. M.; Shaughnessy, K. H.; Lewis, J. P.; Gannett, P. M. *Biophys. Chem.* **2011**, *154*, 41-48.

(45) Perlow, R. A.; Broyde, S. *J. Mol. Biol.* **2001**, *309*, 519-536.

(46) Toth, B.; Gannett, P. *Mycopathologia* **1993**, *124*, 73-77.

(47) Kohda, K.; Tsunomoto, H.; Kasamatsu, T.; Sawamura, F.; Terashima, I.; Shibutani, S. *Chem. Res. Toxicol.* **1997**, *10*, 1351-1358.

(48) Gannett, P. M.; Ye, J. P.; Ding, M.; Powell, J.; Zhang, Y.; Darian, E.; Daft, J.; Shi, X. L. *Chem. Res. Toxicol.* **2000**, *13*, 1020-1027.

(49) Gannett, P. M.; Shi, X. L.; Lawson, T.; Kolar, C.; Toth, B. *Chem. Res. Toxicol.* **1997**, *10*, 1372-1377.

(50) Gannett, P. M.; Lawson, T.; Miller, M.; Thakkar, D. D.; Lord, J. W.; Yau, W. M.; Toth, B. *Chem.-Biol. Interact.* **1996**, *101*, 149-164.

(51) Qi, S. F.; Wang, X. N.; Yang, Z. Z.; Xu, X. H. *Comput. Theor. Chem.* **2011**, *965*, 84-91.

(52) Gannett, P. M.; Powell, J. H.; Rao, R.; Shi, X. L.; Lawson, T.; Kolar, C.; Toth, B. *Chem. Res. Toxicol.* **1999**, *12*, 297-304.

(53) Gannett, P. M.; Heavner, S.; Daft, J. R.; Shaughnessy, K. H.; Epperson, J. D.; Greenbaun, N. L. *Chem. Res. Toxicol.* **2003**, *16*, 1385-1394.

(54) Liu, R.; Liu, H.; Chen, X.; Kirby, M.; Brown, P. O.; Zhao, K. J. *Cell* **2001**, *106*, 309-318.

(55) Jain, N.; Reshetnyak, Y. K.; Gao, L.; Chiarelli, M. P.; Cho, B. P. *Chem. Res. Toxicol.* **2008**, *21*, 445-452.

(56) Liang, F. T.; Meneni, S.; Cho, B. P. *Chem. Res. Toxicol.* **2006**, *19*, 1040-1043.

(57) Mao, B.; Hingerty, B. E.; Broyde, S.; Patel, D. J. *Biochemistry* **1998**, *37*, 81-94.

(58) Patel, D. J.; Mao, B.; Gu, Z. T.; Hingerty, B. E.; Gorin, A.; Basu, A. K.; Broyde, S. *Chem. Res. Toxicol.* **1998**, *11*, 391-407.

(59) Shibutani, S.; Fernandes, A.; Suzuki, N.; Zhou, L.; Johnson, F.; Grollman, A. P. *J. Biol. Chem.* **1999**, *274*, 27433-27438.

(60) Vanhoute, L. P. A.; Westra, J. G.; Retel, J.; Vangrondelle, R. *Carcinogenesis* **1988**, *9*, 1017-1027.

(61) Wang, F.; DeMuro, N. E.; Elmquist, C. E.; Stover, J. S.; Rizzo, C. J.; Stone, M. P. *J. Am. Chem. Soc.* **2006**, *128*, 10085-10095.

(62) Watt, D. L.; Utzat, C. D.; Hilario, P.; Basu, A. K. *Chem. Res. Toxicol.* **2007**, *20*, 1658-1664.

(63) Zhang, L.; Shapiro, R.; Broyde, S. *Chem. Res. Toxicol.* **2005**, *18*, 1347-1363.

(64) Bichara, M.; Fuchs, R. P. P. *J. Mol. Biol.* **1985**, *183*, 341-351.

(65) Brown, K.; Hingerty, B. E.; Guenther, E. A.; Krishnan, V. V.; Broyde, S.; Turteltaub, K. W.; Cosman, M. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 8507-8512.

(66) Burnouf, D. Y.; Wagner, J. E. *J. Mol. Biol.* **2009**, *386*, 951-961.

(67) Gu, Z. T.; Gorin, A.; Hingerty, B. E.; Broyde, S.; Patel, D. J. *Biochemistry* **1999**, *38*, 10855-10870.

(68) Geacintov, N. E.; Broyde, S.; Buterin, T.; Naegeli, H.; Wu, M.; Yan, S. X.; Patel, D. J. *Biopolymers* **2002**, *65*, 202-210.

(69) Cho, B. P.; Zhou, L. *Biochemistry* **1999**, *38*, 7572-7583.

(70) Milhe, C.; Dhalluin, C.; Fuchs, R. P. P.; Lefevre, J. F. *Nucleic Acids Res.* **1994**, *22*, 4646-4652.

(71) Milhe, C.; Fuchs, R. P. P.; Lefevre, J. F. *Eur. J. Biochem.* **1996**, *235*, 120-127.

(72) Garcia, A.; Lambert, I. B.; Fuchs, R. P. P. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 5989-5993.

(73) Fuchs, R. P. P.; Koffel-Schwartz, N.; Pelet, S.; Janel-Bintz, R.; Napolitano, R.; Becherel, O. J.; Broschard, T. H.; Burnouf, D. Y.; Wagner, J. *Biochem. Soc. Trans.* **2001**, *29*, 191-195.

(74) Schorr, S.; Carell, T. *ChemBioChem* **2010**, *11*, 2534-2537.

(75) Shapiro, R.; Ellis, S.; Hingerty, B. E.; Broyde, S. *Chem. Res. Toxicol.* **1998**, *11*, 335-341.

(76) Liang, F. T.; Cho, B. P. *Biochemistry* **2010**, *49*, 259-266.

(77) Cho, B. P.; Beland, F. A.; Marques, M. M. *Biochemistry* **1992**, *31*, 9587-9602.

(78) Nolan, S. J.; McNulty, J. M.; Krishnasamy, R.; McGregor, W. G.; Basu, A. K. *Biochemistry* **1999**, *38*, 14056-14062.

(79) Mao, B.; Vyas, R. R.; Hingerty, B. E.; Broyde, S.; Basu, A. K.; Patel, D. J. *Biochemistry* **1996**, *35*, 12659-12670.

(80) Wang, F.; Elmquist, C. E.; Stover, J. S.; Rizzo, C. J.; Stone, M. P. *Biochemistry* **2007**, *46*, 8498-8516.

(81) Elmquist, C. E.; Wang, F.; Stover, J. S.; Stone, M. P.; Rizzo, C. J. *Chem. Res. Toxicol.* **2007**, *20*, 445-454.

(82) Wu, X. Y.; Shapiro, R.; Broyde, S. *Chem. Res. Toxicol.* **1999**, *12*, 895-905.

(83) Belguisevalladier, P.; Fuchs, R. P. P. *J. Mol. Biol.* **1995**, *249*, 903-913.

(84) Cho, B. P.; Beland, F. A.; Marques, M. M. *Biochemistry* **1994**, *33*, 1373-1384.

(85) Eckel, L. M.; Krugh, T. R. *Biochemistry* **1994**, *33*, 13611-13624.

(86) Eckel, L. M.; Krugh, T. R. *Nat. Struct. Biol.* **1994**, *1*, 89-94.

(87) Mao, B.; Hingerty, B. E.; Broyde, S.; Patel, D. J. *Biochemistry* **1998**, *37*, 95-106.

(88) Liang, F. T.; Cho, B. P. *Chem. Res. Toxicol.* **2011**, *24*, 597-605.

(89) Jia, L.; Shafirovich, V.; Shapiro, R.; Geacintov, N. E.; Broyde, S. *Biochemistry* **2006**, *45*, 6644-6655.

(90) Gannett, P. M.; Sura, T. P. *Chem. Res. Toxicol.* **1993**, *6*, 690-700.

(91) Koehl, P.; Valladier, P.; Lefevre, J. F.; Fuchs, R. P. P. *Nucleic Acids Res.* **1989**, *17*, 9531-9541.

(92) Tan, X. Z.; Suzuki, N.; Grollman, A. P.; Shibutani, S. *Biochemistry* **2002**, *41*, 14255-14262.

(93) Jain, N.; Meneni, S.; Jain, V.; Cho, B. P. *Nucleic Acids Res.* **2009**, *37*, 1628-1637.

(94) Lindsley, J. E.; Fuchs, R. P. P. *Biochemistry* **1994**, *33*, 764-772.

(95) Hilario, P.; Yan, S. X.; Hingerty, B. E.; Broyde, S.; Basu, A. K. *J. Biol. Chem.* **2002**, *277*, 45068-45074.

(96) Lakshman, M. K.; Keeler, J. C.; Ngassa, F. N.; Hilmer, J. H.; Pradhan, P.; Zajc, B.; Thomasson, K. A. *J. Am. Chem. Soc.* **2007**, *129*, 68-76.

(97) Broyde, S.; Wang, L. H.; Zhang, L.; Rechkoblit, O.; Geacintov, N. E.; Patel, D. J. *Chem. Res. Toxicol.* **2008**, *21*, 45-52.

(98) Cai, Y.; Patel, D. J.; Geacintov, N. E.; Broyde, S. *J. Mol. Biol.* **2007**, *374*, 292-305.

(99) Christian, T. D.; Romano, L. J. *Biochemistry* **2009**, *48*, 5382-5388.

(100) Dai, Q.; Xu, D. W.; Lim, K.; Harvey, R. G. *J. Org. Chem.* **2007**, *72*, 4856-4863.

(101) Huang, X. W.; Colgate, K. C.; Kolbanovskiy, A.; Amin, S.; Geacintov, N. E. *Chem. Res. Toxicol.* **2002**, *15*, 438-444.

(102) Yang, T. L.; Huang, Y. H.; Cho, B. P. *Chem. Res. Toxicol.* **2006**, *19*, 242-254.

(103) Cavalieri, E. L.; Rogan, E. G. *Pharmacol. Ther.* **1992**, *55*, 183-199.

(104) Cavalieri, E.; Rogan, E. *Environ. Health Perspect.* **1985**, *64*, 69-84.

(105) Cavalieri, E. L.; Rogan, E. G.; Li, K. M.; Todorovic, R.; Ariese, F.; Jankowiak, R.; Grubor, N.; Small, G. J. *Chem. Res. Toxicol.* **2005**, *18*, 976-983.

(106) Chakravarti, D.; Pelling, J. C.; Cavalieri, E. L.; Rogan, E. G. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 10422-10426.

(107) Okahashi, Y.; Iwamoto, T.; Suzuki, N.; Shibutani, S.; Sugiura, S.; Itoh, S.; Nishiwaki, T.; Ueno, S.; Mori, T. *Nucleic Acids Res.* **2010**, *38*.

(108) Akanni, A.; AbulHajj, Y. J. *Chem. Res. Toxicol.* **1997**, *10*, 760-766.

(109) Akanni, A.; Abul-Hajj, Y. J. *Chem. Res. Toxicol.* **1999**, *12*, 1247-1253.

(110) Cavalieri, E. L.; Rogan, E. G. *Future Oncol.* **2010**, *6*, 75-91.

(111) Zahid, M.; Saeed, M.; Ali, M. F.; Rogan, E. G.; Cavalieri, E. L. *Free Radic. Biol. Med.* **2010**, *49*, 392-400.

(112) Brown, K. L.; Voehler, M. W.; Magee, S. M.; Harris, C. M.; Harris, T. M.; Stone, M. P. *J. Am. Chem. Soc.* **2009**, *131*, 16096-16107.

(113) Carvajal, M. In *Food Contaminants: Mycotoxins and Food Allergens*; Siantar, D. P., Trucksess, M. W., Scott, P. M., Herman, E. M., Eds.; Amer Chemical Soc: Washington, 2008; Vol. 1001, p 13-55.

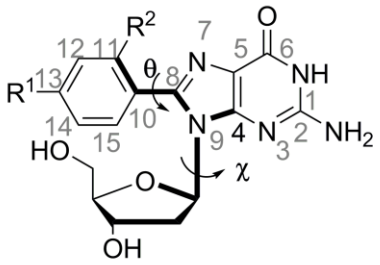
(114) Rogan, E. G.; Cavalieri, E. L.; Tibbels, S. R.; Cremonesi, P.; Warner, C. D.; Nagel, D. L.; Tomer, K. B.; Cerny, R. L.; Gross, M. L. *J. Am. Chem. Soc.* **1988**, *110*, 4023-4029.

(115) Schlitt, K. M.; Sun, K. W. M.; Paugh, R. J.; Millen, A. L.; Navarro-Whyte, L.; Wetmore, S. D.; Manderville, R. A. *J. Org. Chem.* **2009**, *74*, 5793-5802.

2 Chapter 2: Structure and Stability of C-linked PhOH-dG Adducts[a]

2.1 Introduction

As discussed in Chapter 1, the mechanism by which C8-PhOH-dG adducts exert their biological activity is poorly understood, where both adduct-induced conformational changes and abasic site formation have been implicated in the mutagenic mechanism of related mutagenic C8 carbon-bonded adducts. This chapter examines the effects of the bulky phenoxyl group of C-linked C8-PhOH-dG adducts on the adduct structure and stability using nucleobase and nucleoside models, which permit direct comparison to the limited available experimental data. In the current chapter only, the naming scheme depicted in Figure 2.1 will be used due to the large number of adducts studied.



R^1 = H, R^2 = H = **1**
R^1 = OH, R^2 = H = **2a**
R^1 = OCH_3, R^2 = H = **2b**
R^1 = CH_3, R^2 = H = **2c**
R^1 = CN, R^2 = H = **2d**
R^1 = CHO, R^2 = H = **2e**
R^1 = H, R^2 = OH = **3a**
R^1 = H, R^2 = OCH_3 = **3b**
R^1 = H, R^2 = CH_3 = **3c**

Figure 2.1 Structure of aryl adducts considered and identification of dihedral angles χ (∠(O4′C1′N9C4), bold) and θ (∠(C11C10C8N9), bold).

Experimental studies of *o*-PhOH-dG (Figure 2.1, **3a**) have focused on the synthesis and spectral properties of the adduct in attempts to gain information about the structure.[1]

[a] Reprinted in part with permission from (1) Millen, A. L.; McLaughlin, C. K.; Sun, K. M.; Manderville, R. A.; Wetmore, S. D. *J. Phys. Chem. A* **2008**, *112*, 3742-3753. Copyright 2008 American Chemical Society and (2) Schlitt, K. M.; Sun, K. W. M.; Paugh, R. J.; Millen, A. L.; Navarro-Whyte, L.; Wetmore, S. D.; Manderville, R. A. *J. Org. Chem.* **2009**, *74*, 5793-5802. Copyright 2009 American Chemical Society. L.N.-W. performed some calculations for PES scans for adducts 2a-2e, 3b,c.

As discussed in Chapter 1, *o*-PhOH-dG can exist in a planar enol conformation with an intramolecular hydrogen bond between the phenolic OH and the imine N7 (Figure 2.2). Phototautomerization of the planar structure through the ESIPT (excited state intramolecular proton transfer) process to the keto form is recognizable by a large Stokes-shifted fluorescence.[1] Thus, ESIPT, coupled with the solvent dependence of this process, was used previously to study the ground-state structural properties of the nucleobase and nucleoside.[1] By analyzing the UV-visible absorbance and emission spectra for evidence of ESIPT, experimental studies concluded that the *o*-PhOH-dG nucleoside preferentially exists in a twisted conformation in most solvents, although a planar conformation is also present in some aprotic solvents.[1] However, in the absence of the sugar moiety, the nucleobase prefers a planar structure, which suggests that the sugar may induce the twist.[1] Indeed, previous studies have revealed that the sugar moiety affects the preferred nucleobase orientation in a unique class of modified nucleobases known as fleximers.[2] It is important to understand the degree of twist in the nucleobase since this property is known to play a role in the conformational preferences and toxicity of other bulky adducts[3] in larger (DNA duplex) systems, as discussed in Chapter 1. Although *o*-PhOH-dG has been studied experimentally, relatively little evidence exists regarding the structure of *p*-PhOH-dG (Figure 2.1, **2a**). Therefore, the conformational preferences of both adducts require further investigation.

Experimental work has also investigated the mechanism of abasic site formation for a variety of C8-dG adducts, since this is often associated with C8 adduct formation, although unexpected at neutral pH.[4-10] The hydrolysis experiments investigated acid-catalyzed glycosidic bond cleavage for the N7-protonated adducts **1-3c** in conditions with varying pH,[11] and results were extrapolated to physiological pH to predict that phenoxyl adducts are

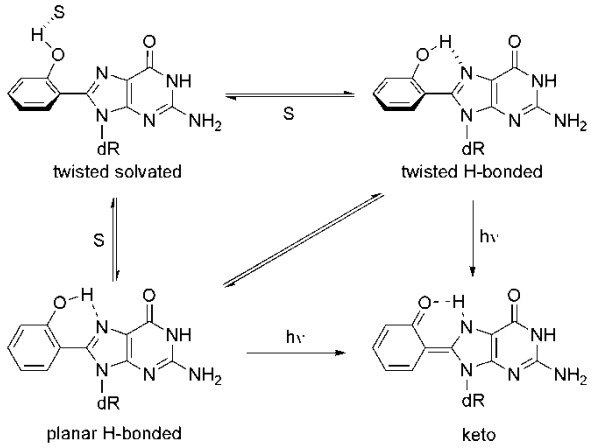


Figure 2.2 Solvent dependence of the intramolecular O–H···N7 hydrogen bond in the *o*-PhOH-dG adduct.

in fact stable in DNA.[11] Experiments considered a range of *para*- (**2a-2e**, Figure 2.1) and *ortho*-phenyl substituents (**3a-3c**, Figure 2.1) of varying electronic and steric properties,[11] since substitution of dG with electron-withdrawing[12-14] C8-substituents greatly accelerates hydrolysis, while electron-donating[15,16] C8-substituents decrease the magnitude of hydrolysis. However, bulky electron-donating[15,17] C8-substituents accelerate hydrolysis compared to the unmodified base despite their electron-donating character. This has been ascribed to release of steric strain upon removal of the deoxyribose moiety,[17,18] where the degree of twist in the nucleoside compared to the nucleobase may affect the stability. Thus, changes in the structures of the nucleoside compared to the nucleobase adducts may explain the relative reactivities of adducts with different phenyl substituents.

The present chapter uses DFT calculations to provide insight into the above experimental results regarding the structure and stability of PhOH-dG adducts. First, the degree of flexibility and preferred conformation of *o*-PhOH-dG, as well as *p*-PhOH-dG (Figure 2.1, **3a** and **2a**, respectively), is thoroughly explored in the current chapter through detailed density functional calculations in the gas phase and in solution to shed light on the experimental data reported above. The corresponding deglycosylated (nucleobase)

analogues of *o*-PhOH-dG and *p*-PhOH-dG are also considered to further evaluate the effects of the sugar moiety on the preferred nucleobase conformation in both adducts. Secondly, to understand the role of steric strain in the relative rates of hydrolysis for adducts **1-3c**, calculations are performed to identify the structures of the neutral and (charged) N7-protonated (N7–H^+) forms of **1-3c** using nucleoside and nucleobase models. Finally, to determine how the C8-aryl moiety contributes to sugar loss, scans of the deglycosylation reactions of neutral and protonated dG and adduct **1** in both the *anti* and *syn* conformations are performed. The computational approach taken in the present chapter reveals fundamental information about the structures and stabilities of the adducts considered, improves interpretation of the experimental results, and provides further insight into the mutagenicity of C8-phenoxyl adducts.

2.2 Computational Details

The B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) potential energy surfaces of the nucleobase models of *o*-PhOH-G and *p*-PhOH-G were considered as a function of θ, the dihedral angle that controls the relative orientation of the phenoxyl and nucleobase rings (Figure 2.1). In *o*-PhOH-G, there exist two possible orientations of the phenol hydroxyl group, namely directed toward (∠(HOC11C10) = 0°) and away from (∠(HOC11C10) = 180°) the nucleobase. Since the relative energies indicate that the lowest energy structures involve the hydroxyl directed toward the nucleobase (See Results and Discussion), only this orientation will be further discussed for the *ortho* nucleoside adduct.

Due to the inherent flexibility within the adducts, the corresponding potential energy surfaces are complicated. Specifically, the sugar puckering and the orientation of the nucleobase about the glycosidic bond (defined by the dihedral angle χ, Figure 2.1) must be considered in addition to rotation about θ. Therefore, systematic B3LYP/6-31G(d) potential energy surface scans were performed where the dihedral angles referred to as χ and θ

(Figure 2.1) were constrained in 10° increments from 0 to 360°. Due to the large number of conformations associated with the sugar moiety and the primary interest in the effects of the sugar moiety on the structure of the nucleobase, preliminary conformational searches were performed using Monte Carlo with MMFF as implemented in the Spartan[19] software package to identify the preferred sugar conformation.[20] The ten lowest energy structures identified from these scans were subsequently optimized with B3LYP/6-31G(d). For both the *para* and *ortho* adducts, the lowest energy conformer involves C2′–*endo* sugar puckering, which is the preferred puckering in B-DNA. Additionally, the C5′ hydroxyl group is directed toward the nucleobase, where further test calculations on B3LYP optimized geometries support that this dihedral is generally preferred over others due to interactions between the C5′ hydroxyl hydrogen and O4′, or the nucleobase. Finally, ∠(HC3′OH) is approximately –60° for all structures.

Based on the results from the conformational searches, three additional selection rules for the geometries of the adducts were used during the B3LYP/6-31G(d) scans. First, the sugar puckering was constrained to be C2′–*endo*, where preliminary studies on the corresponding (*ortho*) damaged imidazole nucleoside revealed that the correct minima are obtained by initially constraining the sugar puckering and subsequently releasing the constraint to yield more accurate energies. Second, although the orientations of the hydroxyl groups were not constrained, only structures with a ∠(HC3′OH) dihedral angle of ~ –60° and the C5′ hydroxyl group directed toward the nucleobase (i.e., a ∠(C4'C5'OH) dihedral angle of approximately 50 – 90°) were considered.

To more accurately determine the geometries and relative energies of the minima and transition states identified from the potential energy surface scans for the adducts, full optimizations (i.e., all constraints released) were subsequently performed. Higher-level (B3LYP/6-311+G(2df,p)) single-point calculations, as well as B3LYP/6-31G(d) frequency

calculations, were performed in the gas phase on all fully optimized structures with scaled (0.9806)[21] zero-point vibrational energies included. Additionally, polarisable continuum model (PCM) single-point calculations were performed on the lowest energy structures in various solvents (cyclohexane, chloroform, acetonitrile, DMSO, and water), which represent a range of polarities for comparison to experimental results.

All minima identified for the *ortho* and *para* adducts were subsequently re-optimized with different R^1 and R^2 substituents (Figure 2.1) to generate a variety of adducts. Additional Monte Carlo calculations were performed on these adducts to ensure that lower energy minima were not missed. Subsequently, each minimum structure was protonated at N7 and re-optimized to study the effects of an acidic environment on the adduct structure and stability. Higher-level (B3LYP/6-311+G(2df,p)) single-point calculations, as well as B3LYP/6-31G(d) frequency calculations, were performed in the gas phase on all fully-optimized structures with scaled (0.9806)[21] zero-point vibrational energies included.

Deglycosylation scans on the various adducts were performed using constrained PCM-B3LYP/6-31G(d) optimizations in the solvent (water) phase for comparison to hydrolysis experiments. In these calculations, only the C1′–N9 bond length was constrained, which was gradually increased in 0.1 Å increments from 1.4 to 3.4 Å and optimized at each step. All calculations were performed using Gaussian 03.[22]

2.3 Results and Discussion

2.3.1 *o*-PhOH-dG Adduct

2.3.1.1 Nucleoside Model

Figure 2.3 (left) displays the results from the B3LYP potential energy surface scan for *o*-PhOH-dG. Close atomic contacts make it difficult to find stable structures in the white regions of the graphs. It is anticipated that the corresponding high-energy structures do not

play a significant role, and therefore the discussion will focus on the more stable regions of the surface.[23]

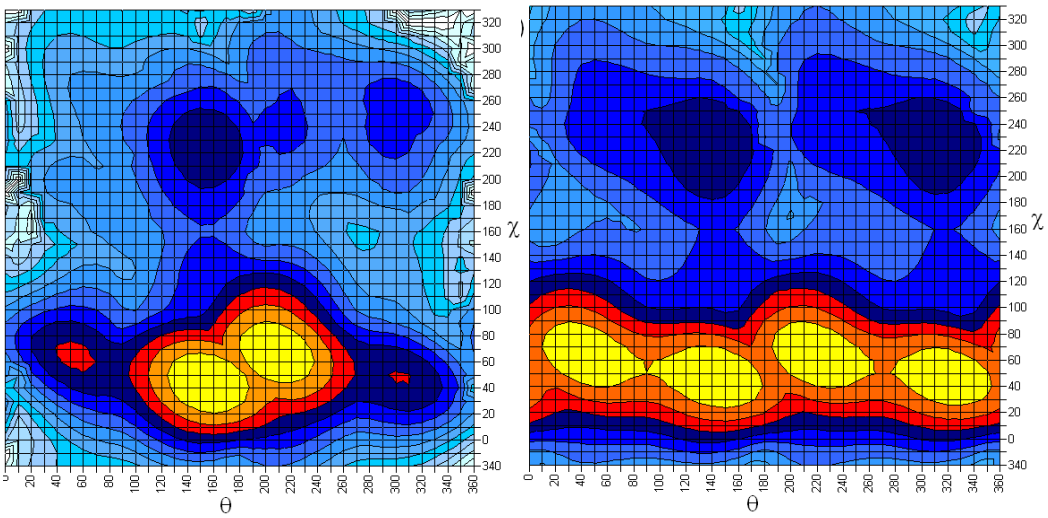


Figure 2.3 B3LYP/6-31G(d) potential energy surface for *o*-PhOH-dG (left) and *p*-PhOH-dG (right) adducts with θ plotted along the x-axis, χ plotted along the y-axis, and energy represented by color, where yellow represents the lowest energy regions and each change in color represents an increase in the relative energy of 10 kJ mol^{-1}.

The contour plot for *o*-PhOH-dG reveals that the global minimum (yellow region) adopts the *syn* conformation. Indeed, there are four *syn* minima (χ ranging from 40 to 70°) that are all lower in energy than any *anti* minima. The *anti* region of the graph displays a less smooth surface in part due to steric crowding as the phenoxyl ring passes over the sugar moiety. However, two *anti* minima can be identified, which are at least 30 kJ mol^{-1} higher in energy than the *syn* global minimum. Rotational barriers between minima can also be estimated from the graph to be approximately 10 – 40 (20 – 30) kJ mol^{-1} for conversion between *syn* (*anti*) minima, and 50 – 80 kJ mol^{-1} for *syn* to *anti* conversions.

Important geometrical parameters, as well as the relative energies, for fully optimized minima and transition states are provided in Figure 2.4. When full optimizations (i.e., all constraints released) are performed on points of interest from the contour plot, the geometries change very little, where χ and θ deviate by less than 5° in minima (16° in

transition states). The sugar puckering also generally remains C2'–*endo*; however, more significant deviations occur in some instances, which will be discussed in more detail below.

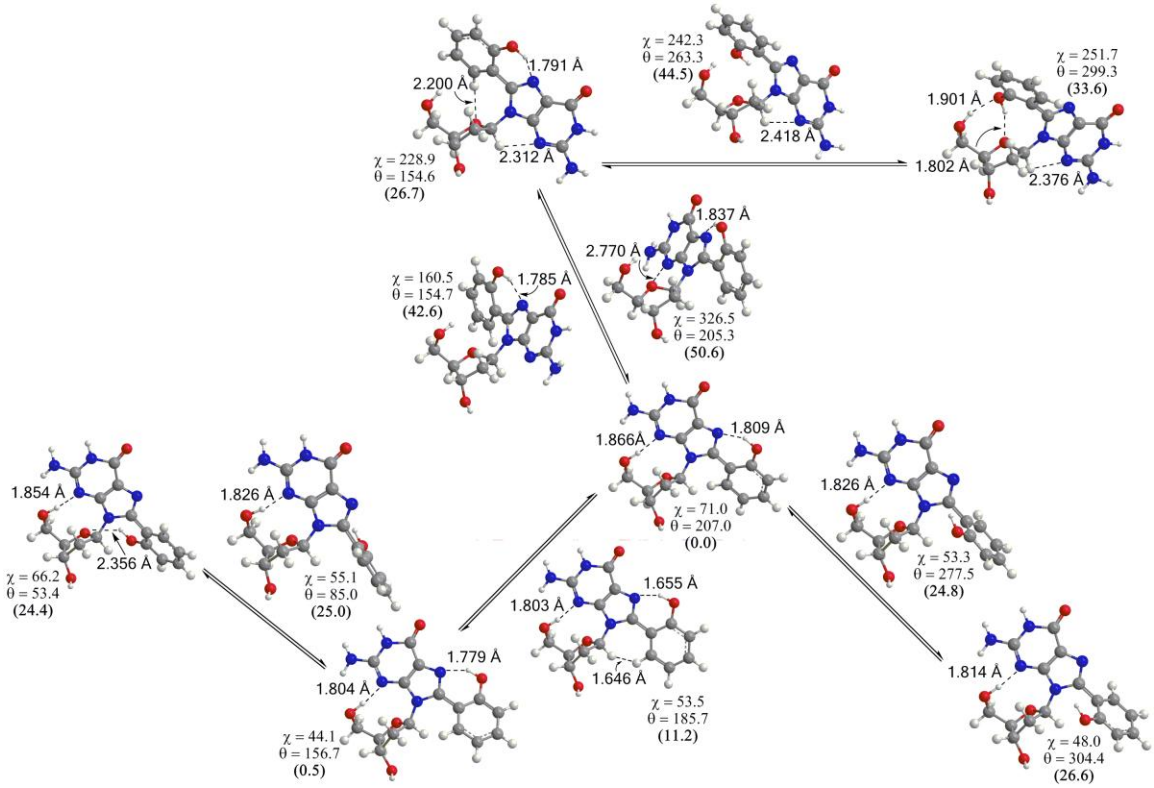


Figure 2.4 Selected bond lengths (Å) and dihedral angles (χ and θ, degrees) for the minima and transition states of the *o*-PhOH-dG adduct fully optimized with B3LYP/ 6-31G(d) (relative energies from B3LYP/6-311+G(2df,p) single-point calculations provided in parenthesis, kJ mol^{-1}).

All *syn* minima contain a strong O5′–H···N3 hydrogen bond (1.80 – 1.87 Å) and the phenoxyl substituent is twisted with respect to the nucleobase, where the magnitude of the twist depends on favorable intramolecular interactions and steric considerations. For example, in the global minimum ($\chi \approx 70°$ and $\theta \approx 205°$), and a nearly thermoneutral local minimum ($\chi \approx 45°$ and $\theta \approx 155°$), the phenoxyl substituent is twisted out of the plane of the nucleobase by approximately 25°. Despite this twisting, both structures contain an O–H···N7 hydrogen bond (1.78 – 1.81 Å), where the phenoxyl hydrogen moves 10 to 13° out of the

plane of the ring to maintain this (albeit slightly twisted) interaction. In the global minimum, the phenoxyl ring is twisted such that a phenol hydrogen is closer to O4′ and steric crowding between hydrogen atoms in the sugar (C1' and C2') and the phenoxyl group is reduced. In two secondary (*syn*) local minima ($\chi \approx 65°$ and $\theta \approx 55°$, and $\chi \approx 50°$ and $\theta \approx 305°$), the phenoxyl ring is twisted by as much as 55° to avoid repulsive interactions between the phenol hydroxyl hydrogen and C1′ of the sugar. The structure with $\theta \approx 55°$ is 2 kJ mol^{-1} more stable than that with $\theta \approx 305°$ due to a weak O–H···O4′ interaction and reduced steric crowding between the C2' and phenoxyl hydrogens.

Although experimental results indicate a possible planar conformation for the adduct, the conformation of *o*-PhOH-dG with a planar nucleobase and strong O–H···N7 hydrogen bond ($\theta \approx 180°$) was found to be a transition state that connects two twisted minima. This is due to repulsive interactions between a phenoxyl C15–H hydrogen and C1′–H, which are reduced by twisting about θ. Nevertheless, the energy barrier for rotation about θ is small (11 – 13 kJ mol^{-1}), which suggests that rotation between the two twisted minima will readily occur and environmental effects could be sufficient to stabilize a (close to) planar structure. The remaining two *syn* transition states contain a perpendicular arrangement of the nucleobase and phenoxyl rings.

Two *anti* local energy minima ($\chi \approx 230 – 250°$) contain interactions between the C5′ hydroxyl and the π-system ($\theta \approx 160°$) or the hydroxyl group ($\theta \approx 300°$) of the phenoxyl substituent. These interactions lead to steric crowding, which is partially alleviated due to changes in the sugar puckering to C1′–*exo*–C2′–*endo* twist. As found for the (*syn*) global minimum, the lowest energy *anti* minimum contains an O–H···N7 hydrogen bond and the phenoxyl ring is twisted by approximately 25° about θ. Nevertheless, the *anti* structures are 25 – 30 kJ mol^{-1} higher in energy than the *syn* structures due to the absence of the strong O5′–H···N3 hydrogen bond. The *anti* minima are connected by a transition state at $\chi \approx 240°$

and θ ≈ 260°, which has C1′–*exo*–C2′–*endo* twist sugar puckering and is approximately 45 kJ mol^{-1} above the *syn* global minima.

Two transition states between the *anti* and *syn* regions of the surface were also characterized (χ ≈ 160° and θ ≈ 155°, and χ ≈ 325° and θ ≈ 205°), which both contain O–H···N7 hydrogen bonds and the phenoxyl hydrogen is located approximately 15° out of the phenyl plane. Although the transition state at χ ≈ 160° has predominately C2′–*endo* sugar puckering, the structure at χ ≈ 325° has a C1′–*exo*–C2′–*endo* twist conformation and is distorted about the glycosidic bond due to repulsion between N3 and O4′. The corresponding barriers (40 – 53 kJ mol^{-1}) indicate that although rotation about χ is more difficult than rotation about θ, the *anti* region of the conformational space will be sampled.

The most important conclusion from the structures discussed thus far is that the lowest energy conformations involve a twisted arrangement of the phenoxyl substituent relative to the nucleobase ring in the gas phase. However, previous experimental studies on *o*-PhOH-dG considered a wide range of solvents and determined a solvent dependence for the preferred structure.[1] Therefore, it is of interest to determine the environmental effects on the results. Since solvation is not expected to provide sufficient stabilization to introduce a preference for the *anti* structures, B3LYP/6-311+G(2df,p) PCM single-point calculations were performed in various solvents (hexane, chloroform, acetonitrile, DMSO, and water) on the four minima and three transition states in the *syn* region of the potential energy surface (Figure 2.5).

Although the energy differences are small, solvation single-point calculations on gas-phase geometries indicate that the potential energy surface changes in different environments. Specifically, the planar transition structure (θ = 180°) is destabilized, and the twisted structures are stabilized, as the polarity of the solvent increases. The calculations

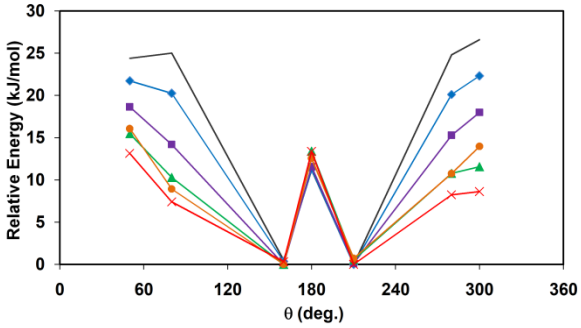


Figure 2.5 B3LYP/6-311+G(2df,p) relative energies of *syn* minima and transition states in gas-phase (black, no symbol), cyclohexane (blue, ♦), chloroform (purple, ■), acetonitrile (green, ▲), DMSO (orange, ●), and water (red, ×) for *o*-PhOH-dG (see Figure 2.4 for geometrical information).

agree with UV-vis spectra obtained for *o*-PhOH-dG in a range of aprotic solvents, as well as the corresponding optical data in water. Specifically, a previous study[24] of the absorbance spectra of *o*-PhOH-dG indicates that both the twisted and planar structures are present. The relative peak heights suggest that the planar conformation is more readily accessible in chloroform than acetonitrile, and that the planar conformation is even less accessible in hexane. In solvents with high polarities (DMSO), the experimental results for *o*-PhOH-dG indicate the presence of a twisted enol structure that cannot phototautomerize[24] (Figure 2.2). Indeed, studies on *o*-PhOH-dG in water also show only one conformation, where it was previously proposed that the O–H···N7 hydrogen bond in *o*-PhOH-dG is broken in hydrogen-bonding solvents such that only a twisted (non-hydrogen bonded) enol structure is present.[1]

Thus, the calculated trend with respect to solvation is in agreement with experiments. Specifically, single-point calculations (Figure 2.5) indicate that the twisted conformer is stabilized in solvents with high polarity (DMSO, water), while the stability of the planar structure increases with a decrease in the dielectric (polarity) of the solvent. Whether the twisted and/or planar nucleobase conformations are observed is likely due to

a balance between the steric effects of the sugar, which influences the twist of the nucleobase, and the strong O–H···N7 hydrogen bond, which anchors the nucleobase in a planar conformation. The computational results imply that as the polarity of the solvent increases, the strength of the hydrogen bond weakens, and therefore the steric interactions become predominant or, in other words, the twisted structure becomes more favourable. This will especially be true in solvents that are capable of forming discrete hydrogen bonds and thereby break the O–H···N7 interaction.

2.3.1.2 Nucleobase Model

To further understand the steric effects of the sugar, the electronic structure of the deglycosylated *o*-PhOH-G was studied as a function of θ (Figure 2.6). As mentioned in the Computational Details, the lowest energy conformations occur when the hydroxyl group is rotated towards the dG moiety (Figure 2.6a) and therefore structures where the hydroxyl group is directed away from dG will not be further discussed (Figure 2.6b). As found for *o*-PhOH-dG, the global minimum for *o*-PhOH-G contains an O–H···N7 hydrogen bond (≈ 1.75 Å). However, although the conformation of *o*-PhOH-dG with a planar nucleobase is a transition state, the global minimum for *o*-PhOH-G is planar (θ = 180°). This supports the proposal that the sugar influences the geometry of the nucleobase and is in agreement with previous findings for other modified nucleobases.[2] Two local minima are approximately 40 kJ mol^{-1} higher in energy, where the phenoxyl group is twisted almost 25° from planarity with respect to the nucleobase to avoid repulsion with the N9 hydrogen. This is less than the 55° twist observed in the presence of the sugar moiety for the higher energy *syn* minima (θ ≈ 55° and 305°), which suggests that the sugar not only introduces twist to the global minimum, but also increases the magnitude of the twist in the local minima.

The transition states connecting the local minima to the global minimum occur when the nucleobase and phenoxyl ring are approximately perpendicular, where the

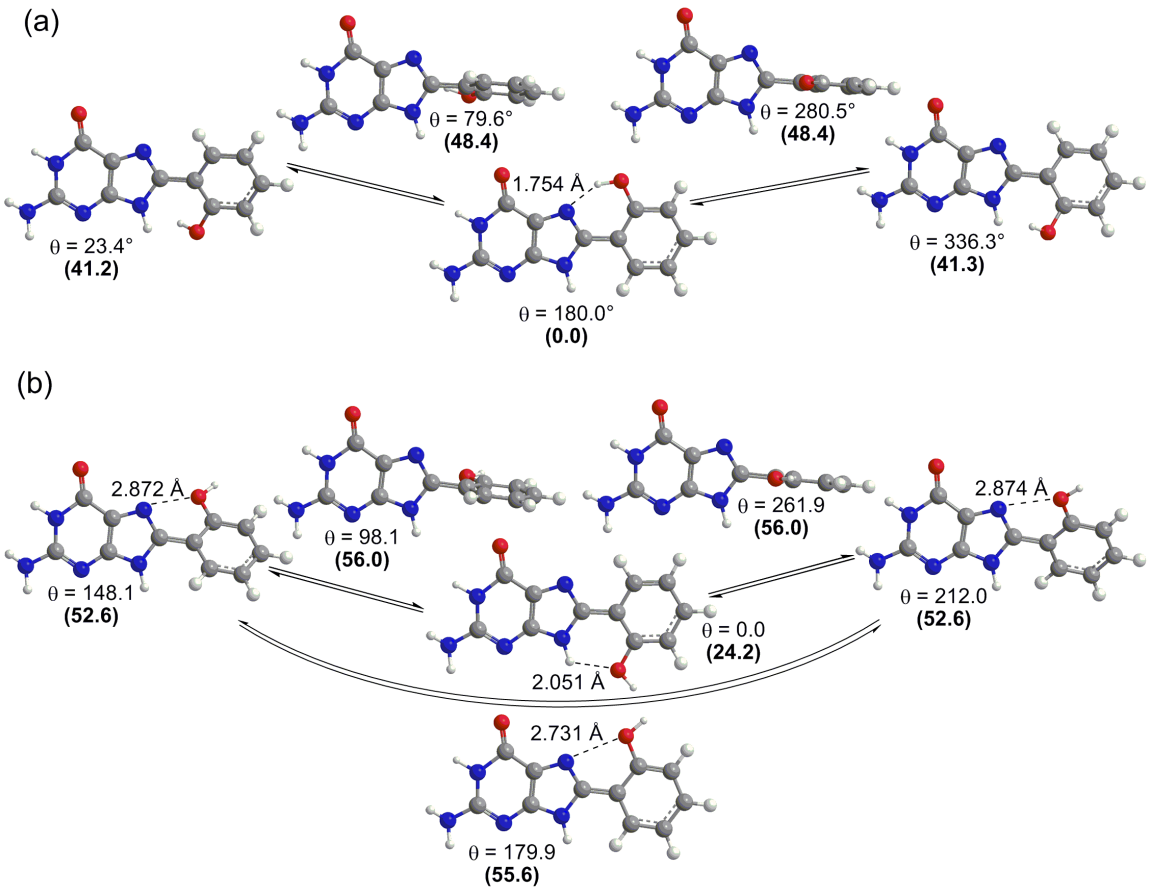


Figure 2.6 Select B3LYP/6-31G(d) bond lengths (Å) and angles (θ, degrees) of *o*-PhOH-G with (a) ∠(HOC11C10) ≈ 0° and (b) ∠(HOC11C10) ≈ 180° (B3LYP/ 6-311+G(2df,p) relative energies from single-point calculations provided in parenthesis, kJ mol^{-1}).

corresponding θ rotational barriers are approximately 48 kJ mol^{-1}. It should be noted that the barrier to rotation is greater for the nucleobase than the deoxynucleoside, which implies that the sugar increases the conformational flexibility of the nucleobase. A transition state connecting the two twisted local minima could not be isolated, which is anticipated to be due to steric crowding between the hydroxyl group and N9 hydrogens.

PCM single-point calculations were performed on the *o*-PhOH-G structures in a variety of solvents (Figure 2.7). As described for *o*-PhOH-dG, the twisted conformations are

relatively more stable in more polar solvents, such as water. However, the planar structure is the global minimum in all solvents for *o*-PhOH-G, including solvents with high polarity. Experimental results in the literature support the conclusions from the calculations. Specifically, the UV-vis spectrum for *o*-PhOH-G in water indicates a planar conformation about θ,[24] and similar results occur in the highly polar solvent DMSO.[1] This contrasts the absorption for *o*-PhOH-dG, which is indicative of a twisted structure.[24] Therefore, experimental observations and calculations both support the proposal that the steric effects of the sugar are responsible for the twist of the nucleobase.

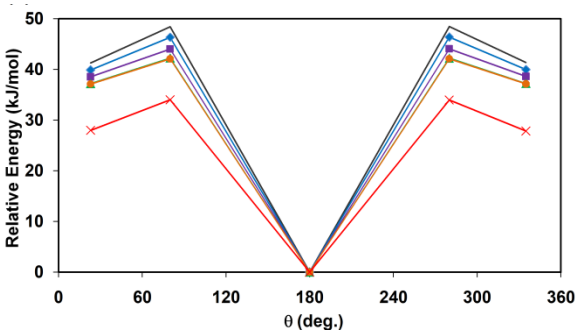


Figure 2.7 B3LYP/6-311+G(2df,p) relative energies of minima and transition states in gas-phase (black, no symbol), cyclohexane (blue, ♦), chloroform (purple, ■), acetonitrile (green, ▲), DMSO (orange, ●), and water (red, ×) for *o*-PhOH-G (see Figure 2.6 for geometrical information).

2.3.2 *p*-PhOH-dG Adduct

2.3.2.1 Nucleoside Model

As discussed in the Introduction, the *p*-PhOH-dG adduct (Figure 2.1, **2a**) was also considered. Investigation of *p*-PhOH-dG will also yield additional structural information about *o*-PhOH-dG. Specifically, the significance of the O–H···N7 hydrogen bond, which is only present in *o*-PhOH-dG, for determining the preferred structure will be revealed. The present section will focus on the structure of *p*-PhOH-dG and the structures of the *o*- and *p*-PhOH-dG adducts will be compared in the following section.

Figure 2.3 (right) displays the B3LYP/6-31G(d) contour plot for rotation about θ and χ in *p*-PhOH-dG. Since many of the distortions found in *o*-PhOH-dG do not occur for *p*-PhOH-dG, the surfaces are greatly simplified, where no high energy structures were left uncharacterized. Similar to the potential energy surface for *o*-PhOH-dG (Figure 2.3, left), the surface for *p*-PhOH-dG has four *syn* minima, which include the global minimum, and two *anti* minima, which are approximately 30 – 40 kJ mol^{-1} higher in energy. In contrast to the *o*-PhOH-dG surface, the *p*-PhOH-dG surface is highly symmetric since rotation about θ by 180° simply changes the orientation of the hydroxyl group on the phenoxyl ring, which does not directly interact with the nucleobase or deoxyribose moiety.

The fully optimized minima and transition states on the potential energy surfaces for *p*-PhOH-dG are displayed in Figure 2.8, where χ, θ and the sugar puckering generally do not deviate significantly from points identified in the contour plots. Closer examination of the structures reveals that all *syn* minima have O5′–H···N3 hydrogen bonds and the structures are twisted about θ by approximately 40°. The two lower energy *p*-PhOH-dG *syn* minima are rotated about θ such that there is an interaction between a phenoxyl hydrogen and O4′. The two planar orientations of the phenoxyl and nucleobase rings are transition structures due to repulsion between hydrogens on the phenoxyl ring and C1′–H, and there are two higher-energy transition states that involve a perpendicular arrangement of the nucleobase and phenoxyl rings. The similarities in the structures of the *p*-PhOH-dG *syn* minima and the lack of strong intramolecular interactions between the phenoxyl ring and the base or sugar moiety leads to very small (less than 3 kJ mol^{-1}) energy differences between minima and low rotational barriers (≈ 5 – 20 kJ mol^{-1}).

The two *anti* minima (χ ≈ 230°) contain interactions between O5'–H and the π–face of the phenoxyl ring and a C1′–H···N3 interaction. The sugar puckering in the *anti* minima is mainly C1′–*exo*-C2′–*endo* twist, which improves the interaction between the C5′ group and

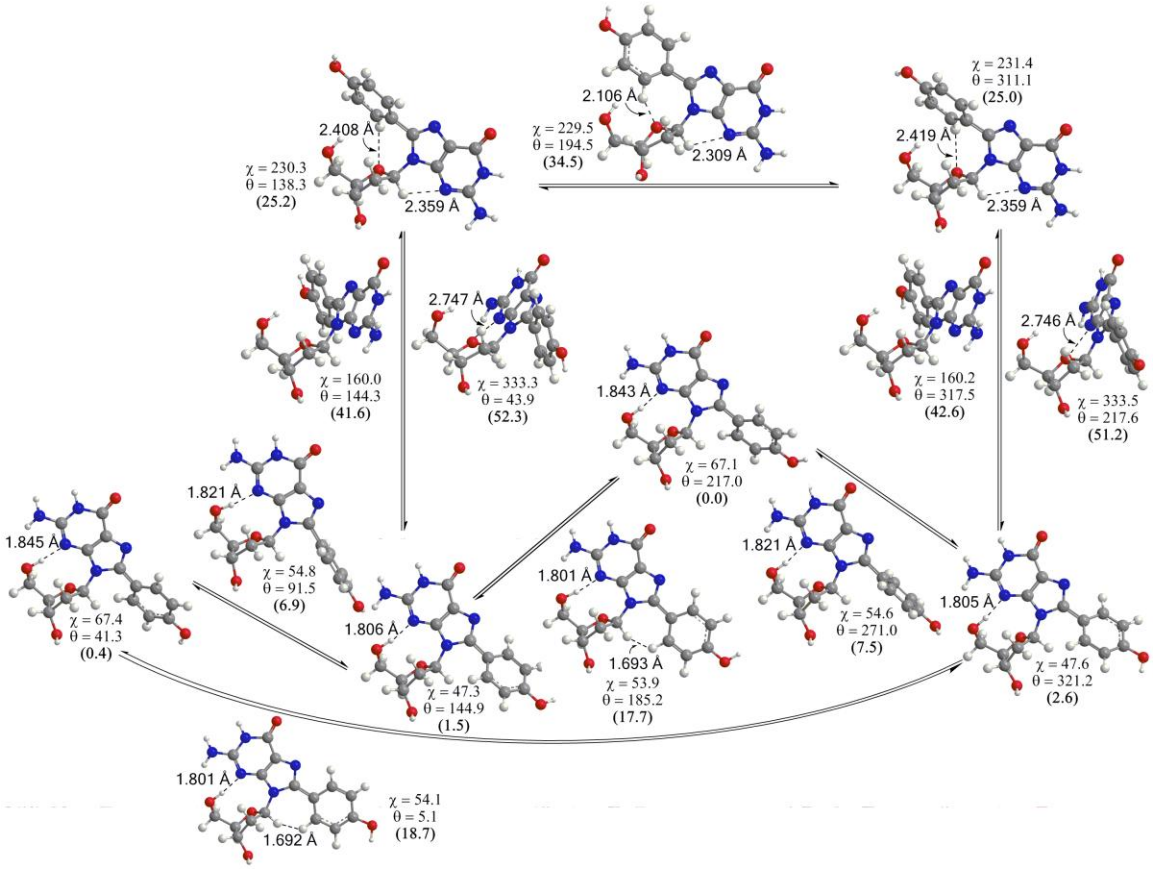


Figure 2.8 Selected bond lengths (Å) and dihedral angles (χ and θ, deg.) for the minima and transition states of the *p*-PhOH-dG adduct fully optimized with B3LYP/ 6-31G(d) (relative energies from B3LYP/6-311+G(2df,p) single-point calculations provided in parenthesis, kJ mol^{-1}).

the phenoxyl ring. The *anti* minima fall approximately 25 kJ mol^{-1} higher in energy than the global minimum, where the barrier for rotation between the *anti* minima is ≈ 10 kJ mol^{-1}. Four transition states connecting the *syn* and *anti* minima were identified, where there is a pair of transition states with χ values of ≈ 160° or 330° that differ in the orientation of the phenol hydroxyl group. The transition states at χ ≈ 330° are 10 kJ mol^{-1} higher in energy due to repulsion between N3 and O4′.

Thus, the calculations predict that the preferred geometry of *p*-PhOH-dG involves a twist about θ. This is supported by UV-vis data for several solvents,[24] which indicate the

presence of a twisted structure, and no absorption that would correspond to a planar conformation.[24] Due to the proposal that the sugar moiety imposes a twisted structure in *o*-PhOH-dG, the structure of *p*-PhOH-dG in the absence of the sugar moiety will now be considered to determine whether the sugar indeed influences the structure.

2.3.2.2 Nucleobase Model

The potential energy surfaces for the deglycosylated version of *p*-PhOH-dG (*p*-PhOH-G) has two planar minima, which simply differ in the orientation of the phenol hydroxyl group (Figure 2.9). The global minimum has the phenoxyl OH directed away from the N9 hydrogen and is only approximately 1 kJ mol^{-1} more stable than the local minimum. The minima are connected through two transition states that contain perpendicular arrangements of the phenoxyl and nucleobase rings. The planar minima identified for the *p*-PhOH-G nucleobase adduct support the conclusion that *o*-PhOH-G in water can adopt a planar conformation in the absence of the (solvated) O–H···N7 hydrogen bond.

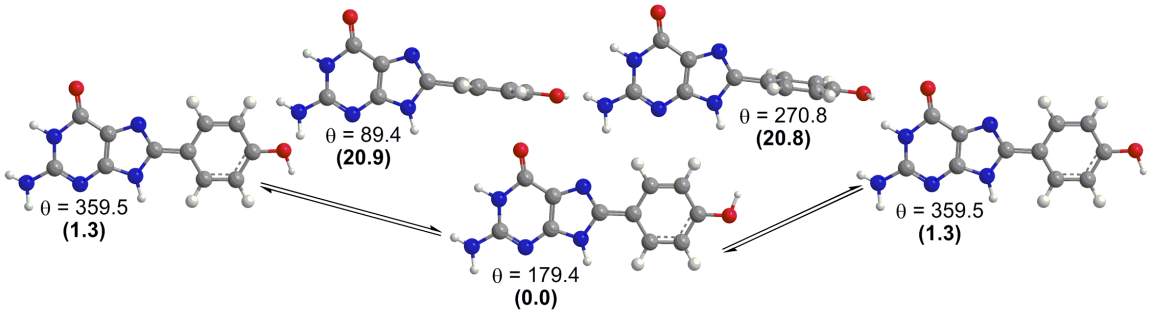


Figure 2.9 B3LYP/6-31G(d) angles (θ, degrees) of *p*-PhOH-G (B3LYP/6-311+G(2df,p) relative energies from single-point calculations provided in parenthesis, kJ mol^{-1}).

Experimental data supports the structural findings.[24] Specifically, the UV-vis spectra of *p*-PhOH-G shows a planar conformation similar to the corresponding *o*PhOH-G adduct. Overall, the study of the *p*-PhOH-dG adduct once again demonstrates the effect of the sugar on the preferred θ orientation, where *p*-PhOH-G is planar, while *p*-PhOH-dG is twisted. The

following section will further discuss the implications of the structure of *p*-PhOH-dG for the observed twist-dependence in *o*-PhOH-dG.

2.3.3 Comparison of *o*- and *p*-PhOH-dG Adducts

Comparison of the potential energy surfaces for the *o*-PhOH-dG and *p*-PhOH-dG adducts reveals important information about their structures. Specifically, calculations agree with experimental results that indicate that the lowest energy conformers are twisted about the carbon-carbon bond connecting the phenoxyl ring to the nucleobase. Calculations show that the *p*-PhOH-dG *syn* minima are twisted about θ by approximately 40°, which is a considerably greater twist than the *syn* global minimum of the *o*-PhOH-dG adduct (25°) (Figure 2.10), but closer to the twist in the two higher energy *p*-PhOH-dG minima (~ 50°). For comparison, the global minima on both the *o*-PhOH-dG and *p*-PhOH-dG surfaces in the absence of the sugar moiety are planar, which suggests that the sugar moiety induces the twist within the nucleobase. The difference in the degree of twist in the *o*- and *p*-PhOH-dG adducts in the presence of the sugar is explained by the competing factor of the O–H···N7 hydrogen bond, where only the lowest energy *o*-PhOH-dG conformers with the smallest θ values have this intramolecular interaction anchoring the nucleobase in a more planar conformation.

Interestingly, the *p*-PhOH-dG global minimum is approximately 20 kJ mol^{-1} higher in energy than the isomeric *o*-PhOH-dG global minimum, which is predominantly due to the loss of the O–H···N7 intramolecular hydrogen bond (Figure 2.10). This difference explains the shallower potential energy surface for the *p*-PhOH-dG adduct (Figure 2.3, left) compared to the *o*-PhOH-dG adduct (Figure 2.3, right). Specifically, the *o*-PhOH-dG adduct minima with θ ranging between 100 and 260° are greatly stabilized relative to the other conformers by the presence of an O–H···N7 hydrogen bond, which is absent in the remaining *o*-PhOH-dG structures (θ ≈ 0 – 100 and 260 – 360°) and all *p*-PhOH-dG conformations. As a result,

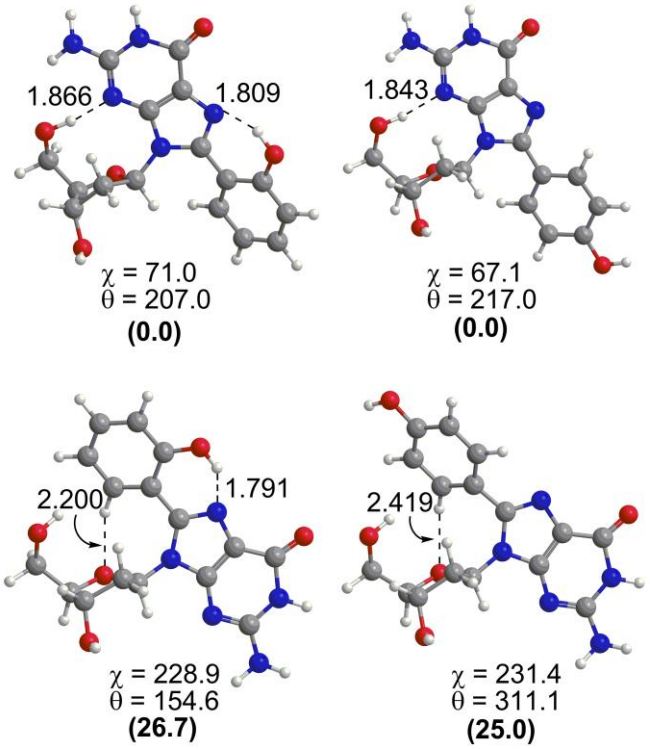


Figure 2.10 Selected bond lengths (Å) and dihedral angles (χ and θ, degrees) for the global *syn* minima (top row) and lowest energy local *anti* minima (bottom row) of the *o*-PhOH-dG (left) and *p*-PhOH-dG (right) adducts fully optimized with B3LYP/ 6-31G(d) (relative energies from B3LYP/6-311+G(2df,p) single-point calculations provided in parenthesis, kJ mol^{-1}).

p-PhOH-dG is much more flexible about θ, where even the transition states for rotation in the *p*-substituted nucleobase adduct in the absence of the sugar are approximately 20 kJ mol^{-1} lower than those for *o*-PhOH-G.

Finally, it is noteworthy that the *syn* structure is the global minimum on the surfaces of *o*-PhOH-dG and *p*-PhOH-dG due to the presence of a strong O5′–H···N3 hydrogen bond. This may have implications for how the adducts interact with the complementary strands within DNA helices. Specifically, the *anti* conformation required for Watson-Crick hydrogen-bonding may be possible in the DNA strand, or the *syn* conformation seen to be preferred in the nucleoside may readily occur, which could lead to mispairing. The implications of the

preferred *syn* structure must be further considered in the following chapters since the C5′ hydroxyl group in the nucleoside is replaced by a phosphodiester backbone in DNA, and therefore the stabilizing (O5′–H···N3) hydrogen bond is not possible.

2.4 Structural Features of Neutral and Protonated C8-Aryl Adducts

The structural features of the neutral and protonated C8 aryl adducts **1-3c** (Figure 2.1) and their deglycosylated analogues were studied in the manner previously presented for the *o*- and *p*-PhOH-dG (**3a** and **2a**) adducts. Due to the similarity in structure of the adducts **1-3c** to the *o*- and *p*-PhOH-dG adducts discussed in detail in the previous section, the results will only be briefly discussed.

As noted for *o*- and *p*-PhOH-dG, *anti* structures are less stable than *syn* structures by ~ 25 kJ mol^{-1} for all neutral C8-aryl-dG adducts **1-3c**. For **1-3c**, all *syn* minima contain a strong O5′–H···N3 hydrogen bond (1.80 – 1.96 Å), and the aryl substituent is twisted with respect to the nucleobase, where the magnitude of the twist angle (θ) depends on steric considerations and favorable intramolecular interactions. For example, the global minima for *p*-adducts **1-2e** are twisted by ~ 37°, while the degree of θ twist increases to 45° for **3c** (R^2 = CH_3) and 55° for **3b** (R^2 = OCH_3) due to steric interactions between R^2 and the nucleobase. In contrast, the *o*-PhOH-dG adduct **3a** (R^2 = OH) global minimum is less twisted (twist = 27°) than any other substituent due to O–H···N7 hydrogen bonding, as discussed in the previous section.

For adducts **1-3b**, the neutral nucleobase global minima, where the sugar is replaced with a hydrogen atom, are planar (θ = 0°), which supports the conclusion that the sugar moiety is inducing the twist within the nucleobase. The barrier to rotation is greater for the nucleobase adducts than the nucleoside adducts, which also supports the proposal that the sugar increases conformational flexibility. The lone exception to this trend is adduct

3c (R^2 = CH_3), where the nucleobase is significantly twisted (by 24°) due to repulsion between the methyl group and sugar.

Representative potential energy surfaces for the N7-protonated adducts (N7–H^+) are shown in Figures 2.11–2.13. For the *p*-substituted adducts **1**-**2e**, the optimized protonated structures are not greatly altered from the neutral structures (Figure 2.11). The global minima are in the same *syn* conformation with a strong O5′–H···N3 hydrogen bond and the *anti* structures are higher in energy (by ~20 kJ mol^{-1}). Table 2.1 reports changes in the twist (θ) angle and C8–C10 bond length upon N7-protonation. The N7-protonated *p*-adducts are more twisted than the neutral species (by about 3 – 6°) due to increased steric interactions between the phenyl ring and the protonated nucleobase (Figure 2.11). Despite the increase in θ, the N7–H^+ nucleosides have shorter C8–C10 bonds, especially when R^1 is electron-donating (OH, OCH_3). The *o*-substituted adducts **3a-c** behave differently upon N7-protonation. Adduct **3a** (R^2 = OH) shows only a small 1.3° increase in θ due to O–H···N7 hydrogen-bonding interactions involving the phenoxyl group in both the neutral and protonated adduct, where the hydroxyl group can be directed toward or away from the dG moiety. In contrast, adduct **3b** (R^2 = OCH_3) becomes much less twisted (by 27.2°) upon N7-protonation due to N7–H···O–CH_3 hydrogen bonding (Figure 2.12), which results in the same degree of twist as neutral **3a**. This preferential stabilization of the N7-protonated adduct for **3b** is also reflected in the relatively large decrease in the C8–C10 bond length. For **3c** (R^2 = CH_3), the N7-protonated adduct is significantly more twisted (22.9°) than the neutral species due to increased sterics (Figure 2.13).

2.5 Stability of C8-Aryl Adducts

Experimental results for the stability of the C8-aryl adducts reveal that the C8-aryl substituent clearly increases the magnitude of the rate constant relative to dG.[11] The effects for the *p*-adducts **1-2e** are ~ 90- to 200-fold larger than for dG, while the effects for the

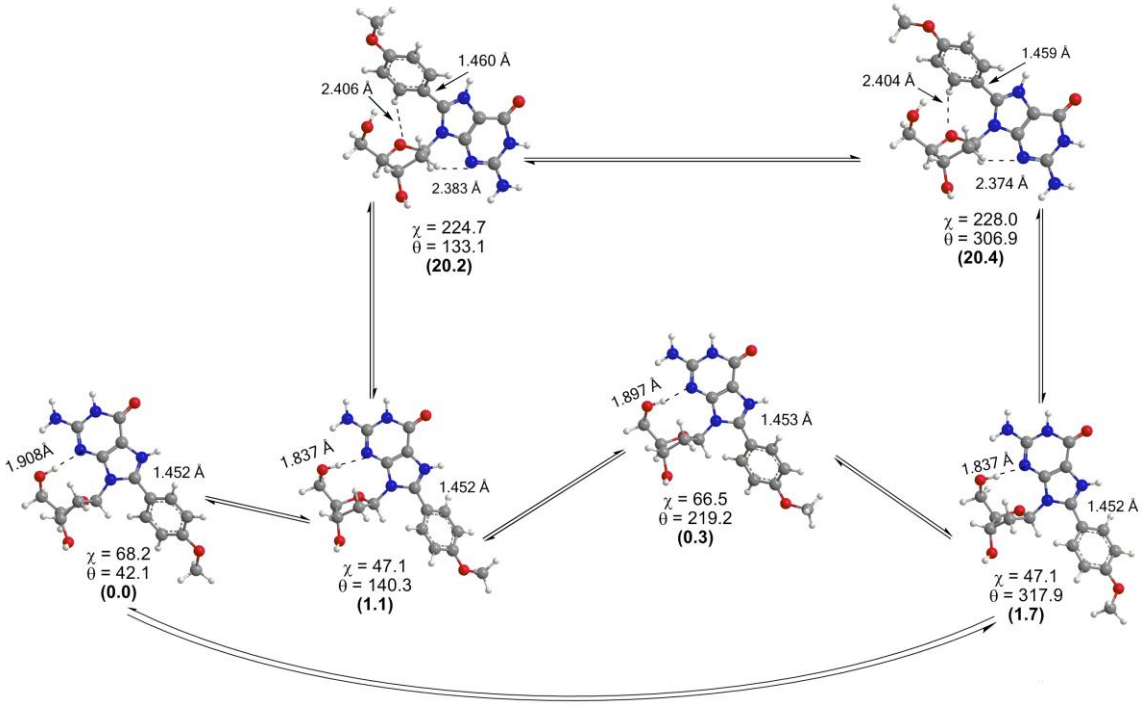


Figure 2.11 Selected bond lengths (Å) and dihedral angles (χ and θ, degrees) for the minima of the N7–H$^+$ adduct 2b **fully optimized with B3LYP/6-31G(d) (relative energies from B3LYP/6-311+G(2df,p) single-point calculations provided in parentheses, kJ mol^{-1}).**

o-adducts **3a-c** are only ~ 9- to 60-fold larger.[11] For comparison, C8-arylamino substituents increase the rate constant by ~ 3- to 15-fold,[17] which is similar to the rate enhancement exerted by the least reactive C8-aryl-dG adduct **3c** ($R^2 = CH_3$).

The differences in the rate of hydrolysis for the *o*-adducts **3a-c** compared to the *p*-adducts **1-2e** is interesting, given that a bulky substituent at C8 is expected to increase the rate of hydrolysis due to the relief of strain.[18] A rationale for the decrease in rates of hydrolysis for **3a-c** relative to **1-2e** is provided by considering the changes in the twist angle (θ) and C8–C10 bond length (Table 2.2) when going from the N7-protonated nucleoside adduct to the neutral nucleobase with N7–H that lacks the sugar moiety (Figure 2.14). For the *p*-adducts **1-2e**, a decrease in twist angle of ~ 40° is observed upon removal of the sugar group from the twisted N7-protonated species to form the planar nucleobase.

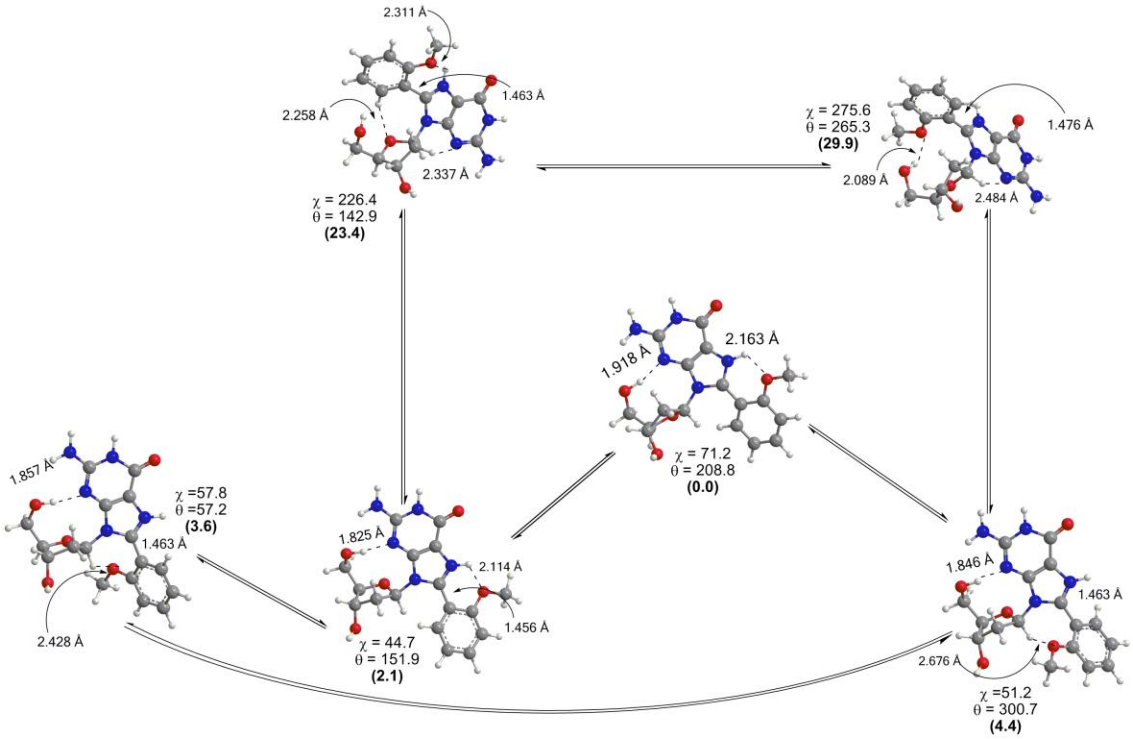


Figure 2.12 Selected bond lengths (Å) and dihedral angles (χ and θ, degrees) for the minima of N7–H⁺ adduct 3b **fully optimized with B3LYP/6-31G(d) (relative energies from B3LYP/6-311+G(2df,p) single-point calculations provided in parentheses, kJ mol⁻¹).**

For **2d** and **2e**, which bear electron-withdrawing groups, the C8–C10 bond length also shortens, suggesting preferential stabilization of the neutral nucleobase compared to the protonated nucleoside. This phenomenon is also expected to increase the rate of hydrolysis. For *o*-adducts **3a** and **3b**, the decrease in twist angle upon sugar removal to form the planar nucleobase is only ~ 28°, because the N7-protonated nucleoside is already relatively planar due to hydrogen-bonding interactions between the R^2 substituent and N7–H⁺. Thus, the relief in strain for **3a** and **3b** upon sugar removal is not as great as for **1-2e**, and hence the rate of hydrolysis is diminished. Differentiation between **3a** and **3b** is apparent from the C8–C10 bond length changes, where the bond becomes shorter for **3a** and longer for **3b**. This suggests preferential stabilization of the neutral nucleobase for **3a**, and hence its rate

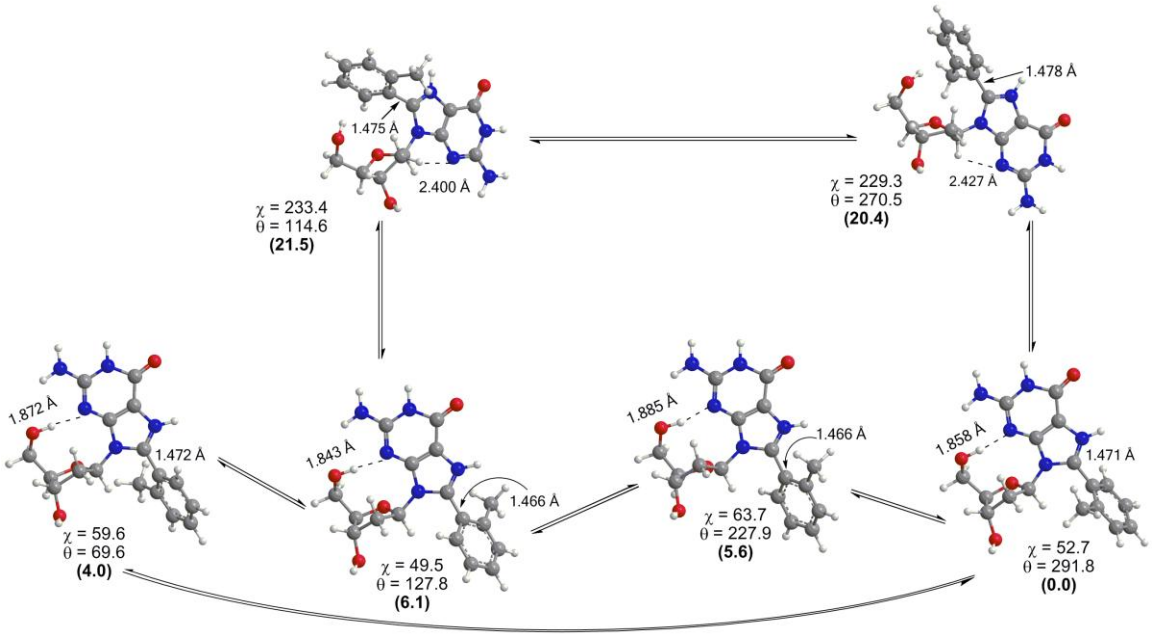


Figure 2.13 Selected bond lengths (Å) and dihedral angles (χ and θ, degrees) for the minima of N7–H⁺ adduct 3c **fully optimized with B3LYP/6-31G(d) (relative energies from B3LYP/6-311+G(2df,p) single-point calculations provided in parentheses, kJ mol^{-1}).**

Table 2.1 Structural features for adducts 1-3c protonated at N7 (N7–H⁺)[a]

Adduct	θ (N7–H⁺)	Δθ[b]	C8–C10 (N7–H⁺)	Δ C8–C10[c]
1 (R^1=H)	42.3	4.8	1.462	–0.011
2a (R^1=OH)	42.8	5.8	1.454	–0.017
2b (R^1=OCH_3)	42.1	4.7	1.453	–0.018
2c (R^1=CH_3)	40.7	3.5	1.459	–0.013
2d (R^1=CN)	42.0	4.0	1.464	–0.006
2e (R^1=CHO)	42.3	4.2	1.464	–0.007
3a (R^2=OH)	28.3	1.3	1.457	–0.007
3b (R^2=OCH_3)	28.8	–26.5	1.456	–0.021
3c (R^2=CH_3)	68.2	22.9	1.471	–0.006

[a] Structures were fully optimized with B3LYP/6-31G(d) (relative energies from B3LYP/6-311+G(2df,p) single-point calculations). [b] Δθ = θ (N^7H^+) – θ (neutral adduct). [c] Δ C8–C10 = C8–C10 (N7–H+) – C8–C10 (neutral adduct).

of hydrolysis is greater than the rate for **3b**. For **3c**, which bears the *o*-CH_3 substituent, the change in twist of 39° suggests considerable relief of strain. However, unlike the other C8-aryl-G adducts, the neutral nucleobase for **3c** is not planar, but has a twist angle θ of 29°

Table 2.2 Structural changes[a] for hydrolysis of adducts 1-3c.

Adduct	Δθ[b]	Δ C8–C10[c]
1 (R^1=H)	–42.3	0.002
2a (R^1=OH)	–42.6	0.009
2b (R^1=OCH_3)	–42.0	0.010
2c (R^1=CH_3)	–39.8	0.004
2d (R^1=CN)	–42.0	–0.004
2e (R^1=CHO)	–42.1	–0.004
3a (R^2=OH)	–28.3	–0.002
3b (R^2=OCH_3)	–28.8	0.010
3c (R^2=CH_3)	–39.2	–0.003

[a] Structures were fully optimized with B3LYP/6-31G(d) (relative energies from B3LYP/ 6-311+G(2df,p) single-point calculations). [b] Δθ = θ ($N7–H^+$ adduct) – θ (neutral nucleobase with N7–H). [c] Δ C8–C10 = C8–C10 ($N7H^+$ adduct) – C8–C10 (neutral nucleobase).

(Figure 2.14). This suggests that the nucleobase is also sterically hindered and that the relief of strain upon sugar removal for **3c** is diminished due to hindrance in the free nucleobase. These results explain the observed trend in experimental rates of acid-catalyzed hydrolysis.

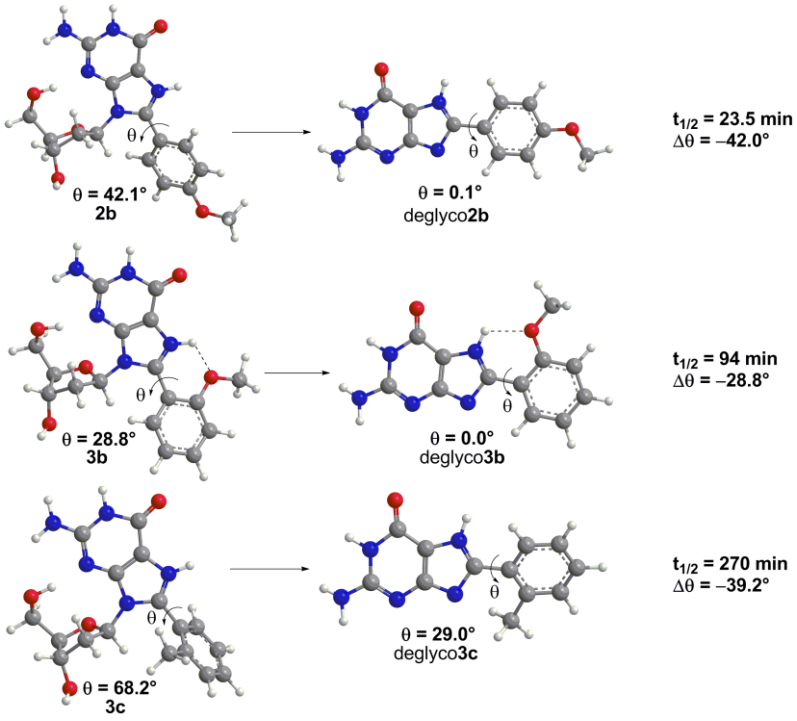


Figure 2.14 Structural changes in the nucleobase upon deglycosylation for 2b, 3b**, and** 3c**.**

Understanding of the effects of the structural changes in the modified base on the rates of deglycosylation can be improved through calculations of the barrier to deglycosylation (Figure 2.15). The structural features that account for the experimentally observed decrease in barrier height upon adduct formation are provided by inspection of the calculated deglycosylation profile for dG vs. **1** shown in Figure 15. By estimating the barriers from the energetic plateaus, it can be seen that although natural dG prefers the *anti* conformation over *syn* by ~11 kJ mol^{-1}, the barrier to deglycosylation of the *syn* conformer (~175 kJ mol^{-1}) is ~20 kJ mol^{-1} less than the corresponding barrier for *anti* (~195 kJ mol^{-1}). Therefore, the *syn* conformation induced by the bulky C8-substituent may contribute to the acceleration of the hydrolysis rates of C8-aryl adducts. Nevertheless, when the barrier of neutral *syn* dG is compared to neutral Ph-dG (**1**), a large effect (~ 40 kJ mol^{-1}) of the bulky C8-phenyl group is still observed (green vs. orange line in Figure 2.15), which suggests C8-substitution may contribute to abasic-site formation. Interestingly, N7-protonation has a larger (80 kJ mol^{-1}, purple vs. red line) effect on the barrier for dG deglycosylation than it does for Ph-dG (50 kJ mol^{-1}, orange vs. light blue line). This diminishes the role of the C8-phenyl substituent in acidic media where N7 is protonated, although the calculations correctly predict that protonated **1** still has the lowest barrier for glycosidic bond cleavage overall (by ~20 kJ mol^{-1}). This role of pH in the relative rates of deglycosylation of modified dG bases versus natural dG is mirrored in experimental results where the rate of hydrolysis for adduct **1** is 20 times greater than dG in 0.1 M HCl and ~ 100 times greater at pH 4.[11] The same phenomenon was observed for the hydrolysis of C8-arylamine adducts that exhibit considerably higher rates of hydrolysis than dG at ~pH 3–7, but became more comparable at pH < 2.[17] Overall, the results suggest a decrease in stability upon adduct formation. However, the barrier to deglycosylation remains relatively high under neutral conditions.

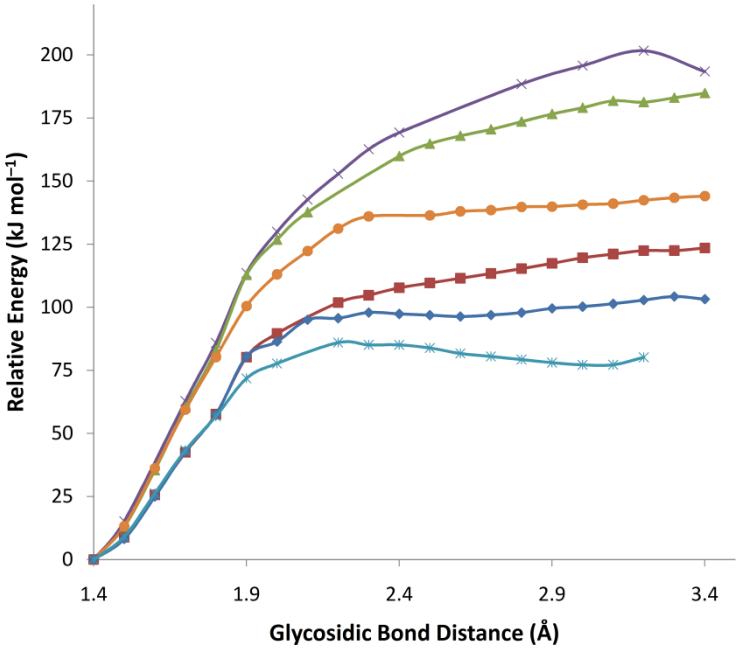


Figure 2.15 The deglycosylation profile calculated with constrained PCM-B3LYP/ 6-31G(d) optimizations for *anti* (×, purple) and *syn* (▲, green) neutral dG, *anti* (■, red) and *syn* (♦, dark blue) N7–H$^+$ dG, and neutral (●, orange) and N7–H$^+$ (*, light blue) *syn* 1.

In fact, experimental hydrolysis rates[11] have been extrapolated to physiological pH (pH = 7) for Ph-dG (**1**),[11] which results in a half-life of $t_{1/2}$ = 25 days. For comparison, a rate for spontaneous loss of purines from duplex DNA at pH 7.4 at 37 °C is ~ 3 x 10^{-11} s^{-1} ($t_{1/2}$ = 730 years).[25] Thus, while C8-aryl-dG adducts are significantly more reactive than dG towards hydrolysis, they are reasonably stable at physiological pH, as supported by the relative barriers to deglycosylation for the neutral adduct in Figure 2.15. Furthermore, they should be even more stable within duplex DNA where purines are more resistant to hydrolysis.[26] The hydrolysis kinetics for C8-aryl-dG adducts **1-3c** suggest that these adducts are unlikely to be intermediates prior to loss of the deoxyribose sugar, as suggested in previous literature, and that an alternative explanation may be required to account for the observation of abasic site formation during C8 aryl adduct formation at physiological pH.

For example, it has been proposed[11] that the radical intermediate formed after aryl radical attack at C8 prior to H-atom abstraction may be responsible for this observation.

2.6 Conclusions

The computational results presented in this chapter demonstrate that the structures and properties of *o*-PhOH-dG (**3a**) and *p*-PhOH-dG (**2a**) do not differ significantly. For both *o*- and *p*-PhOH-dG damage, the global minimum has a *syn* conformation about the glycosidic bond due to the presence of an O5′–H···N3 hydrogen bond, and the *anti* minimum is approximately 20 – 30 kJ mol^{-1} higher in energy. Although both types of damaged nucleosides have a preferred twisted structure about the carbon–carbon bond connecting the phenoxyl and the nucleobase rings, the nucleobase adducts are planar. This suggests that the presence of the sugar introduces a twisted structure. Nevertheless, the degree of the twist is less significant in the *o*-PhOH-dG global minimum compared to *p*-PhOH-dG due to the formation of an O–H···N7 hydrogen bond, which helps anchor the planar orientation. Calculated barriers between different conformers of the nucleosides and nucleobases suggest that the sugar also increases the nucleobase flexibility by lowering rotational barriers. This is significant because twist in related adducts is associated with greater disruptions to the structure of the DNA helix.[3]

Studies of the structure and stability of neutral and protonated adducts **1-3c** complement and aid the interpretation of experimental results regarding the acid-catalyzed hydrolysis of the adducts. Specifically, calculations agree with experimental conclusions[11] that the C8-aryl-dG nucleoside adducts are more prone to acid-catalyzed hydrolysis than dG. Furthermore, experiments show that C8-aryl-dG adducts **1-2e** that bear *para*-substituents have significantly larger rate constants for deglycosylation than for dG, while the effects for the *ortho*-adducts **3a-c** are not as enhanced.[11] Calculations support the hypothesis that relief of steric strain upon removal of the deoxyribose sugar moiety

provides a rationale for this relative reactivity, where the degree of twist induced by the steric clash between the bulky group and the sugar or nucleobase moiety determine the extent of steric relief upon loss of the sugar. Furthermore, deglycosylation pathways show that protonation at N7, the bulky substituent, and the *syn* conformation all contribute to the observed increase in the rate of C1′–N9 glycosidic bond cleavage. Despite the enhanced reactivity of these bases in acid compared to dG, experiments predict they are relatively stable at physiological pH.[11] Although abasic site formation is known to accompany C8-aryl-dG adduct formation, the computational results support the hypothesis that these nucleoside adducts are unlikely intermediates leading to abasic site formation at physiological pH, and other mechanisms may be responsible for this observation.

In summary, the computational approach outlined in this chapter has provided greater knowledge of the structure of the *o*-PhOH-dG and *p*-PhOH-dG C-bonded adducts, as well as a better understanding of the likely stability of these adducts. However, the implications of these findings within the context of DNA have not yet been fully explored. For example, if the adducts are stable and the *syn* conformer is preferred, there could be a potential for mispairing in the DNA strand. Therefore, the next chapter investigates the hydrogen-bonding base pair preferences of the adducts in both the *anti* and *syn* conformations to determine the potential for mismatch stabilization.

2.7 References

(1) McLaughlin, C. K.; Lantero, D. R.; Manderville, R. A. *J. Phys. Chem. A* **2006**, *110*, 6224-6230.

(2) Bardon, A. B.; Wetmore, S. D. *J. Phys. Chem. A* **2005**, *109*, 262-272.

(3) Broyde, S.; Wang, L. H.; Zhang, L.; Rechkoblit, O.; Geacintov, N. E.; Patel, D. J. *Chem. Res. Toxicol.* **2008**, *21*, 45-52.

(4) Rogan, E. G.; Cavalieri, E. L.; Tibbels, S. R.; Cremonesi, P.; Warner, C. D.; Nagel, D. L.; Tomer, K. B.; Cerny, R. L.; Gross, M. L. *J. Am. Chem. Soc.* **1988**, *110*, 4023-4029.

(5) Cavalieri, E. L.; Rogan, E. G.; Li, K. M.; Todorovic, R.; Ariese, F.; Jankowiak, R.; Grubor, N.; Small, G. J. *Chem. Res. Toxicol.* **2005**, *18*, 976-983.

(6) Hiramoto, K.; Kaku, M.; Sueyoshi, A.; Fujise, M.; Kikugawa, K. *Chem. Res. Toxicol.* **1995**, *8*, 356-362.

(7) Lawson, T.; Gannett, P. M.; Yau, W. M.; Dalal, N. S.; Toth, B. *J. Agric. Food Chem.* **1995**, *43*, 2627-2635.

(8) Gannett, P. M.; Lawson, T.; Miller, M.; Thakkar, D. D.; Lord, J. W.; Yau, W. M.; Toth, B. *Chem.-Biol. Interact.* **1996**, *101*, 149-164.

(9) Kohda, K.; Tsunomoto, H.; Kasamatsu, T.; Sawamura, F.; Terashima, I.; Shibutani, S. *Chem. Res. Toxicol.* **1997**, *10*, 1351-1358.

(10) Gannett, P. M.; Powell, J. H.; Rao, R.; Shi, X. L.; Lawson, T.; Kolar, C.; Toth, B. *Chem. Res. Toxicol.* **1999**, *12*, 297-304.

(11) Schlitt, K. M.; Sun, K. W. M.; Paugh, R. J.; Millen, A. L.; Navarro-Whyte, L.; Wetmore, S. D.; Manderville, R. A. *J. Org. Chem.* **2009**, *74*, 5793-5802.

(12) Tretyakova, N. Y.; Burney, S.; Pamir, B.; Wishnok, J. S.; Dedon, P. C.; Wogan, G. N.; Tannenbaum, S. R. *Mutat. Res.-Fundam. Mol. Mech. Mutagen.* **2000**, *447*, 287-303.

(13) Yermilov, V.; Rubio, J.; Ohshima, H. *FEBS Lett.* **1995**, *376*, 207-210.

(14) Laayoun, A.; Decout, J. L.; Lhomme, J. *Tetrahedron Lett.* **1994**, *35*, 4989-4990.

(15) Jordan, F.; Niv, H. *Nucleic Acids Res.* **1977**, *4*, 697-709.

(16) Garrett, E. R.; Mehta, P. J. *J. Am. Chem. Soc.* **1972**, *94*, 8532-&.

(17) Novak, M.; Ruenz, M.; Kazerani, S.; Toth, K.; Nguyen, T. M.; Heinrich, B. *J. Org. Chem.* **2002**, *67*, 2303-2308.

(18) Hovinen, J.; Glemarec, C.; Sandstrom, A.; Sund, C.; Chattopadhyaya, J. *Tetrahedron* **1991**, *47*, 4693-4708.

(19) Spartan '02 Wavefunction, Inc. Irvine, CA

(20) Systematic searches were also performed on the ortho deoxyguanosine adduct, which resulted in the same global minimum as the Monte Carlo search. Consequently, only Monte Carlo searches were performed for the remaining adducts.

(21) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502-16513.

(22) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03, Revisions C.02 and D.01; Gaussian, Inc.: Wallingford, CT, 2004.

(23) A small number of points throughout the main part of the surfaces could also not be optimized under geometrical constraints; however, since realistic structures for the 6 to 8 points surrounding the missing structures could be easily obtained, the energies for the problematic structures were extrapolated to produce a smooth curve.

(24) Millen, A. L.; McLaughlin, C. K.; Sun, K. M.; Manderville, R. A.; Wetmore, S. D. *J. Phys. Chem. A* **2008**, *112*, 3742-3753.

(25) Lindahl, T.; Karlstro.O *Biochemistry* **1973**, *12*, 5151-5154.

(26) Gates, K. S.; Nooner, T.; Dutta, S. *Chem. Res. Toxicol.* **2004**, *17*, 839-856.

3 Chapter 3: Watson-Crick Versus Hoogsteen Base-Pairing for C-Linked PhOH-dG Adducts[a]

3.1 Introduction

The preceding chapter discussed the structure and stability of the *o*- and *p*-PhOH-dG adducts, where the *syn* conformation was found to be favoured over the native *anti* conformation. However, since environmental factors (hydrogen bonding, stacking) have been shown to affect the conformational preferences of other C8 adducts,[1,2] it is important to consider base-pairing preferences for both *anti* and *syn* conformations to determine which conformer is preferentially stabilized by hydrogen-bonding interactions. This will also lead to a better understanding of the role of conformational heterogeneity in the mutagenicity of the adducts.

As discussed in Chapter 1, the Watson-Crick face is available for hydrogen bonding to the neighboring strand in the helix when the adduct adopts the *anti* conformation (Figure 3.1, left), while only the Hoogsteen face of the adduct is accessible for hydrogen-bond pair formation when the adduct adopts the *syn* conformation (Figure 3.1, right). Furthermore, it has been experimentally observed that the addition of a bulky substituent to C8 of the purines enhances Hoogsteen interactions due to the presence of additional hydrogen-bonding sites.[3,4] Therefore, bulky C8 adducts can have preferred hydrogen-bonding patterns that are inaccessible to natural DNA, which may change the structural and base-pairing preferences of the adducts compared to natural bases. Specifically, if one conformation of the bulky adduct forms a relatively strong hydrogen-bonded complex with a natural base, then this particular base pair may be preferentially incorporated into the DNA helix upon replication, where the adduct will adopt the corresponding conformation.

[a] Reprinted in part with permission from Millen, A. L.; Churchill, C. D. M.; Manderville, R. A.; Wetmore, S. D. *J. Phys. Chem. B* **2010**, *114*, 12995-13004. Copyright 2010 American Chemical Society. C.D.M.C. performed some preliminary calculations.

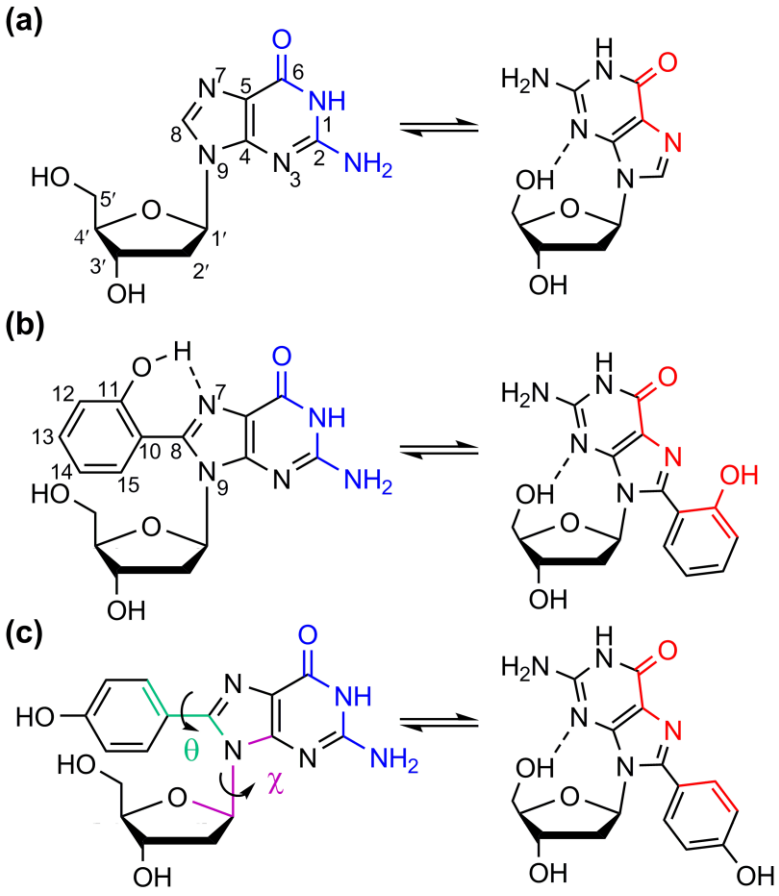


Figure 3.1 The Watson-Crick (blue) and Hoogsteen (red) bonding faces of (a) natural dG, and the (b) *o*- and (c) *p*-PhOH-dG adducts considered in the present work.

This chapter evaluates the Watson-Crick and Hoogsteen binding strengths between the *o*- or *p*-PhOH-dG adduct and the four natural DNA bases. It is important to consider both the *o*- and *p*-PhOH-dG adducts since their structural differences may lead to changes in the (Hoogsteen) hydrogen-bonding preference. Therefore, each adduct may adopt different conformations and exhibit unique mutagenic properties. In the absence of experimental data regarding the base-pairing preferences of these adducts, the results in this chapter will help determine the potential of hydrogen-bonding interactions to stabilize the *syn* conformation. Additionally, since previous studies have successfully used calculated binding strengths between damaged bases and the four natural nucleobases to predict the most

probable DNA mutations,[5-8] this work represents the first step in modeling the potential mutagenic profile of phenoxyl adducts, as well as highlighting systems of interest for further study in a duplex environment.

3.2 Computational Details

The Watson-Crick and Hoogsteen hydrogen-bonded complexes formed between the *anti* or *syn* conformations of the *o*- or *p*-PhOH-dG adducts and the natural bases were optimized in the gas phase with B3LYP/6-31G(d). This level of theory has been found to provide accurate geometries for hydrogen-bonded complexes involving the natural nucleobases.[9] Two computational models were used for the adduct base pairs: 1) the nucleobase model, where the DNA sugar moiety was replaced with a hydrogen atom for both the adduct and the natural bases, and 2) the nucleoside model, where a deoxyribose unit was attached to the damaged base and a methyl group replaced the sugar moiety in the natural nucleoside. In the adduct nucleoside, the ∠(C4′C5′OH) dihedral angle was constrained to 180° to more closely represent a DNA duplex environment, where the C5′ hydroxyl is a phosphate group. When the natural nucleoside is paired with the adduct, the sugar attached to the natural base is reduced to a methyl group since the structures of the natural bases are not greatly affected by the presence of the sugar.[10] Inclusion of both the sugar and the methyl group allows the R(C1′···C1′) distance, which dictates the width of the helix, and the ∠(N9C1′C1′) angle, which dictates the base-pair opening angle with respect to N9 in the adduct, to be measured and compared to average experimental values for B-DNA double helices (10.6 Å and 55°, respectively).[11] Although the final optimized geometries may deviate from those expected for B-DNA, all initial guesses were built to have R(C1′···C1′) distances and ∠(N9C1′C1′) angles appropriate for B-DNA. In some cases, this involved multiple starting orientations, where the resulting lowest energy structure was selected for detailed analysis. For comparison, identical calculations were performed using

the nucleoside model, where the adduct was replaced with natural dG. Only the nucleoside model was used for the natural guanine complexes since this model best allows for a structural comparison of natural and damaged hydrogen-bonded complexes (see Results and Discussion for further details). B3LYP/6-311+G(2df,p) single-point calculations were performed to obtain accurate binding energies, which also include counterpoise corrections to account for the basis set superposition error (BSSE).[12] In all cases, the magnitude of the BSSE correction was less than 3 kJ mol^{-1}.

The interaction energies were calculated as the difference between the energy of the optimized complex and sum of the energies of the monomers in the geometry adopted in the complex (E_{Int}). Since the monomers in this orientation are not fully optimized, the resulting binding energies (E_{Int}) do not include zero-point vibrational energy (ZPVE) corrections. Subsequently, the deformation energy (E_{Def}) of the monomers upon complexation was included in the binding strengths, such that $E_{Tot} = E_{Int} + E_{Def}$. In other words, the total interaction energies (E_{Tot}) were calculated as the difference between the energy of the optimized complex and the sum of the energies of the individually optimized monomers. To allow more direct comparisons of E_{Int} and E_{Tot}, the binding energies calculated using the individually optimized monomers (E_{Tot}) discussed in the main text do not include ZPVE corrections. Nevertheless, the values of E_{Tot} that include scaled (0.9806)[13] ZPVE corrections are provided in the data tables (in parentheses) for comparison, where the corrections range between 2.3 and 6.3 kJ mol^{-1}.

All calculations were performed using Gaussian 03.[14]

3.3 Results and Discussion

3.3.1 Natural dG

A great deal of computational work has considered the strengths and binding arrangements between the natural nucleobases.[8,10,15-41] However, these studies used a

variety of bases, base-pairing modes, computational models, and levels of theory. Therefore, before discussing the hydrogen-bonding patterns of the modified bases, it is necessary to systematically examine the structure and stability of the particular hydrogen-bonded complexes between dG and each of the four natural (N1- or N9-methylated) DNA nucleosides of interest in this work (Figure 3.2). This information will subsequently be used to illustrate the differences in behaviour between natural and damaged dG. For example, it is important to consider whether the interaction energies of the natural pairs are weakened or strengthened by the bulky group. Additionally, this data will help determine whether inclusion of the bulky group permits new hydrogen-bonding patterns. For example, Hoogsteen complexes between the natural bases generally do not meet the structural criteria for an optimal fit in the DNA helix,[42] but interactions with bulky substituents may significantly alter the geometries of these base pairs and thereby produce more viable structures.

First, the binding strengths of base pairs between the Watson-Crick face (*anti* conformation about the glycosidic bond (Figure 3.1a, left)) of dG and the neighboring DNA strand will be discussed. The natural dG:C Watson-Crick pair contains three strong hydrogen bonds (Figure 3.2a) and has a calculated (E_{Int}) interaction strength of –113 kJ mol^{-1} (Table 3.1), which is in good agreement with previous literature.[16-18,23] In contrast, the dG mismatched pairs (pairs not involving C) with (*anti*) T, (*syn*) A or (*syn*) G contain only two hydrogen bonds, which are similar to previously reported structures,[18,21,25,30,36,39] and the corresponding interaction strengths range from –49 to –62 kJ mol^{-1} (Table 3.1). The structural features of the natural dG:C base pair (R(C1′···C1′) = 10.7 Å and ∠(N9C1′C1′) = 52.9°) are similar to those expected for DNA, whereas those for the three mismatched pairs show some deviations (Figure 3.2a). Additionally, propeller-twist and buckle distortions result in a highly non-planar dG:G base pair. Therefore, despite the magnitude of the

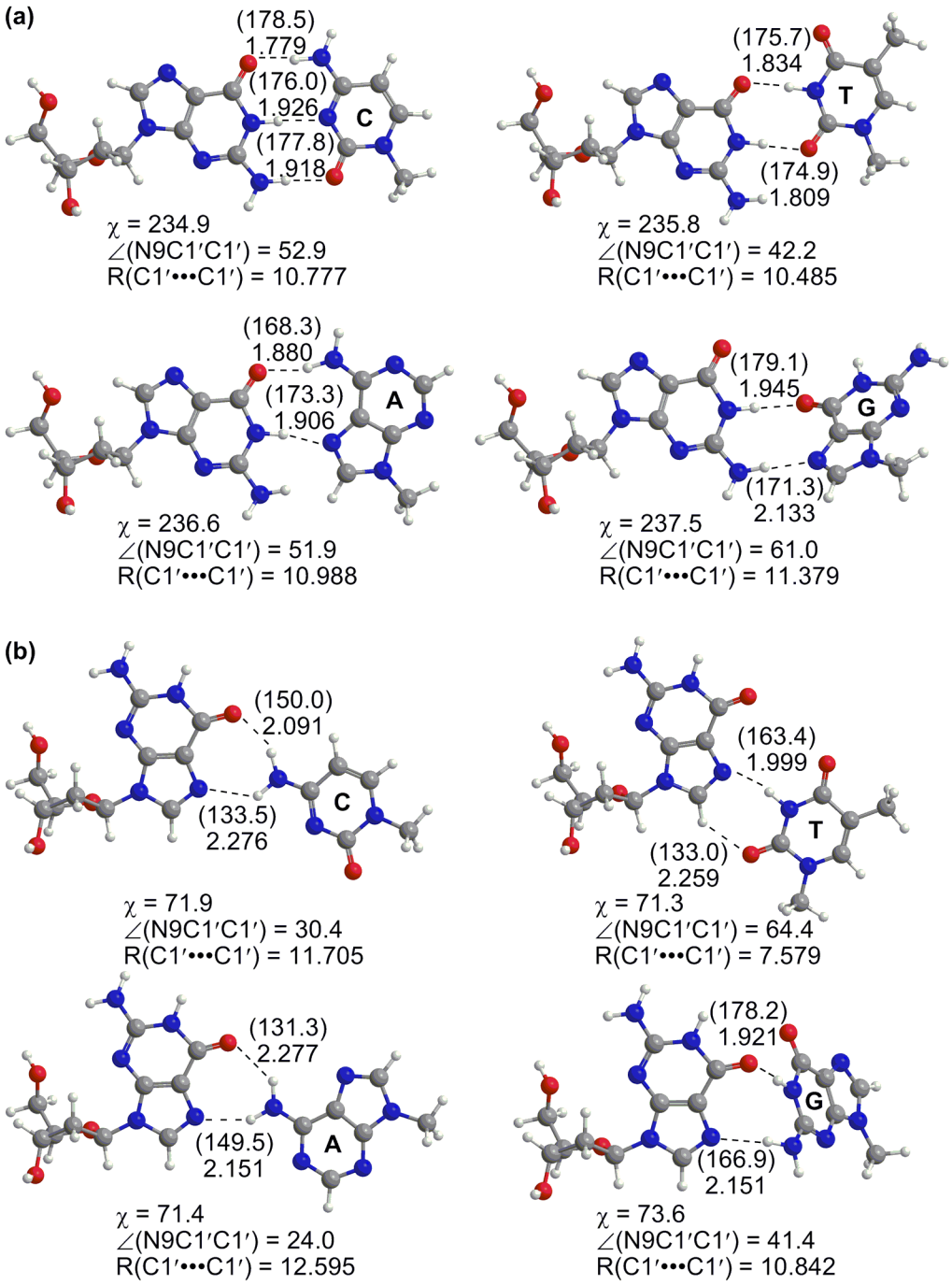


Figure 3.2 B3LYP/6-31G(d) structures (distances in Å, angles in °) for complexes between the (a) Watson-Crick (*anti*) and (b) Hoogsteen (*syn*) faces of dG and each of the four natural DNA nucleoside models.

calculated binding strengths of (fully-optimized) isolated mismatches, their low stability relative to the C pair and their less suitable structure support the observed high specificity

of the Watson-Crick face of dG for hydrogen bonding with C. This exemplifies the importance of considering the interplay between hydrogen-bonding strength and structure in order to predict the binding preference of nucleobases, where small models risk overestimation of binding strengths compared to base pairs in DNA, since free optimization allows more stable bonding arrangements to form.

Table 3.1 B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) binding strengths (kJ mol^{-1}) between the Watson-Crick (*anti*) or Hoogsteen (*syn*) face of natural dG and the four natural DNA nucleoside models.

	Cytosine		Thymine	
	WC	**Hoogsteen**	**WC**	**Hoogsteen**
E_{Int}[a]	−113.5	−37.8	−61.5	−26.4
E_{Tot}[b]	−100.3 (−93.8)	−36.0 (−32.3)	−53.1 (−49.1)	−24.1 (−21.1)
E_{Def}	13.1	1.8	8.5	2.3
Select Previously Published Values	105.9,[c] 120.7,[d] 107.7,[e] 112.0,[f] 116.7,[g] 99.7[h]	39.8[e]	63.3,[e] 66.6,[g] 58.2[h]	
	Adenine		**Guanine**	
	WC	**Hoogsteen**	**WC**	**Hoogsteen**
E_{Int}[a]	−61.8	−28.8	−48.6	−53.2
E_{Tot}[b]	−53.5 (−47.6)	−27.7 (−24.6)	−46.6 (−42.1)	−51.2 (−46.1)
E_{Def}	8.3	1.1	2.0	2.0
Select Previously Published Values	66.6,[g] 54.9[h]		50.2[i]	

[a] The binding strength calculated as the energy difference between the complex and the monomers in the complex geometry (excluding a ZPVE correction). [b] The binding strength calculated as the energy difference between the complex and the individually optimized monomers (excluding a ZPVE correction). Binding energies including (0.9806) scaled ZPVE corrections are provided in parentheses. [c] See Reference 16. [d] See Reference 17. [e] See Reference 18. [f] See Reference 23. [g] See Reference 30. [h] See Reference 36. [i] See Reference 39.

When dG rotates about the glycosidic bond to adopt the *syn* conformation, the Hoogsteen-bonding face becomes available for base-pair formation (Figure 3.1a, right), which has two hydrogen-bond acceptors (O6 and N7). Similar to previously published results,[18] the dG:C complex is significantly destabilized (by 75 kJ mol^{-1}, Table 3.1) upon hydrogen bonding with the Hoogsteen face of dG compared to the Watson-Crick face, and is

a poor structural match for the natural DNA helix (Figure 3.2b). Interestingly, the strongest (E_{Int}) interaction involving the Hoogsteen face of dG occurs with (*anti*) G (–53 kJ mol^{-1}), followed by C (–38 kJ mol^{-1}), (*anti*) A (–29 kJ mol^{-1}) and T (–26 kJ mol^{-1}). Although all of these base pairs deviate from the optimal structure of pairs found in DNA, the structure of the dG:G complex is the closest to B-DNA, which supports previous findings that this Hoogsteen pair can be misincorporated.[43]

The structure of the dG monomer does not significantly change upon binding with the natural bases. Therefore, when the monomer deformation energy is taken into account by calculating the (E_{Tot}) binding strengths as the energy difference between the complexes and the individually optimized monomers, all Watson-Crick and Hoogsteen binding strengths decrease by less than 13 kJ mol^{-1} and the conclusions drawn from the data hold. Since structural modifications to the natural base may alter the specificity of dG to hydrogen bond with C, the following section examines the hydrogen-bonding preferences of *o*-PhOH-dG.

3.3.2 *o*-PhOH-dG Adduct

3.3.2.1 Nucleobase Model

The discussion of the hydrogen-bonded complexes of *o*-PhOH-dG begins by considering the nucleobase model (in the absence of the deoxyribose moiety). In Chapter 2, the structure of the nucleobase monomer was discussed in detail, where the lowest-energy structure of the nucleobase was determined to be planar with respect to the θ dihedral angle (θ = 180°, Figure 3.1). This planarity results in a strong intramolecular hydrogen bond between the hydroxyl group of the phenoxyl ring and N7 of the guanine nucleobase (Figure 3.1b). Although other higher-energy conformations of the adduct exist that are twisted with respect to θ, the lowest-energy (planar) conformation was used to generate initial guesses for the hydrogen-bonded complexes with the natural nucleobases. It is anticipated that the

planar nucleobase is especially relevant to the DNA helix where it may lead to more favorable stacking. Nevertheless, as discussed below, the adduct does not always adopt a planar geometry in fully-optimized hydrogen-bonded complexes.

In all Watson-Crick pairs involving (*anti*) *o*-PhOH-G (Figure 3.3a), the damaged base remains planar (θ = 180°). Since the Watson-Crick bonding faces of *o*-PhOH-G and natural G are identical, the adduct nucleobase forms hydrogen bonds to C with a strength (E_{Int} = −118 kJ mol^{-1}, Table 2) and structure (Figure 3.3a) similar to dG (−113 kJ mol^{-1} (Table 3.1) and Figure 3.2a). Although the adduct base pair involving (*anti*) T is planar (Figure 3.3a, top), the mismatched pairs involving the (*syn*) purines exhibit propeller-twist and buckle distortions (Figure 3.3a, bottom), as discussed for the corresponding natural pairs. The (E_{Int}) binding strengths of the three mismatched pairs range from −55 to −65 kJ mol^{-1} (Table 3.2). Despite being significant in magnitude, these binding arrangements are only half as stable as the adduct:C base pair. Furthermore, there is little difference between the strength of pairs involving the adduct (Table 3.2) or the natural base (Table 3.1). These results for the nucleobase model indicate that the addition of the bulky group to C8 does not affect the Watson-Crick hydrogen-bonding pattern of G and therefore the Watson-Crick face of the adduct will be as selective in base pairing as natural dG.

Similar to natural guanine, the Hoogsteen face of (*syn*) *o*-PhOH-G has two hydrogen-bond acceptors (O6 and N7, Figure 3.1b). However, the hydroxyl group on the phenyl ring provides a third hydrogen-bonding site, which can act as a hydrogen-bond donor or acceptor depending on the hydroxyl orientation. The direct participation of the bulky substituent in hydrogen bonding results in a Hoogsteen pair with C (Figure 3.3b) that is stabilized (by 64 kJ mol^{-1}) relative to the corresponding pair with natural dG. Furthermore, the Hoogsteen face of the adduct paired with C is only slightly less stable (by 16 kJ mol^{-1}) than the corresponding Watson-Crick pair (Table 3.2). In comparison, the difference in the

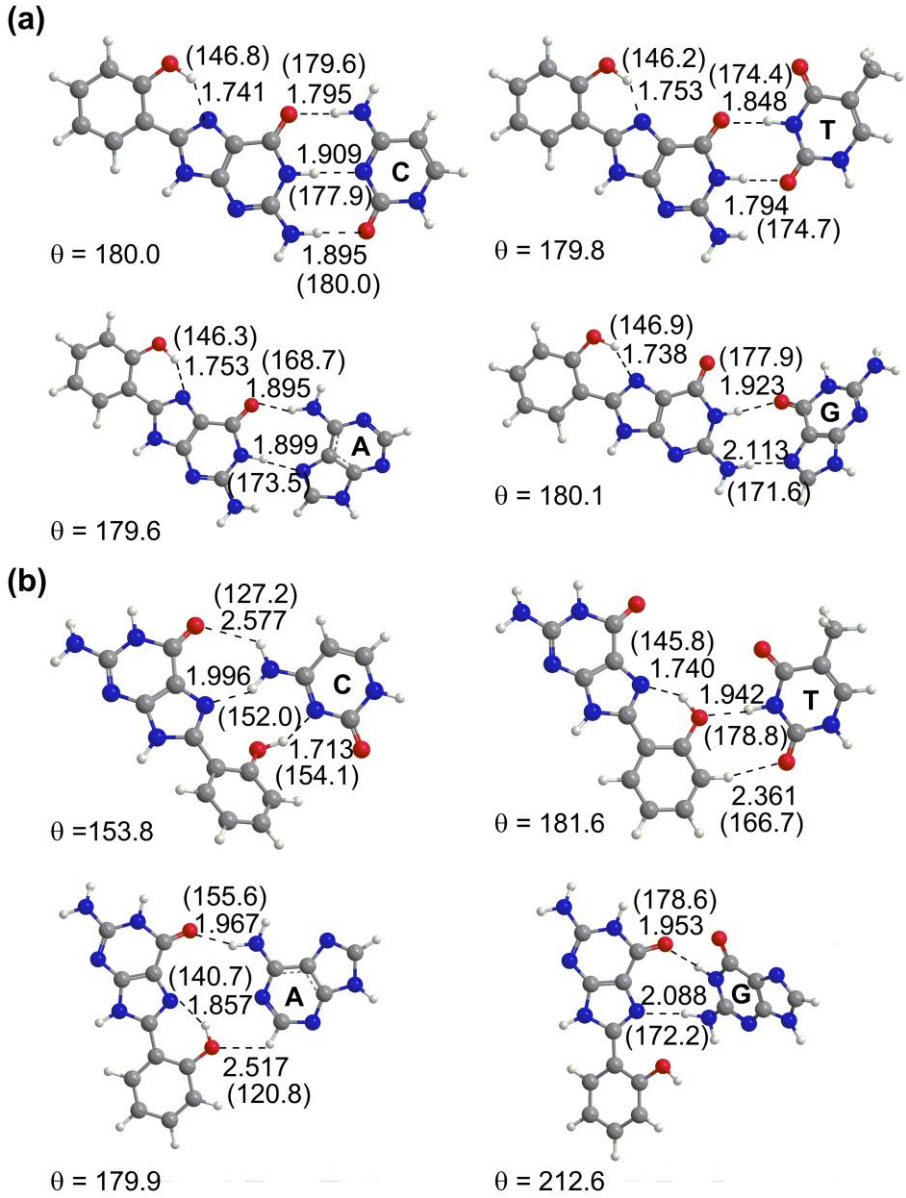


Figure 3.3 B3LYP/6-31G(d) structures (distances in Å, angles in degrees) for complexes between the (a) Watson-Crick (*anti*) and (b) Hoogsteen (*syn*) faces of the *o*-PhOH-G (nucleobase) adduct and each of the four natural DNA nucleobases.

strengths of these two C hydrogen-bonding patterns with natural dG is 75 kJ mol^{-1}. The comparable stabilities of the adduct:C Hoogsteen and Watson-Crick interactions indicates that hydrogen bonding with C in DNA will not significantly destabilize the *syn* conformation of the adduct relative to the *anti* conformation. Therefore, unlike natural dG, the presence of the bulky C8 group may afford two favorable conformations of the adduct when paired with

Table 3.2 B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) binding strengths (kJ mol^{-1}) between the Watson-Crick (*anti*) or Hoogsteen (*syn*) face of the *o*- or *p*-PhOH-G (nucleobase) adduct and the four natural DNA nucleobases.

		Cytosine		Thymine	
		WC	**Hoogsteen**	**WC**	**Hoogsteen**
$\mathbf{E_{Int}}$[a]	*ortho*	–118.1	–101.8	–64.8	–16.3
	para	–114.9	–57.4	–64.0	–[b]
$\mathbf{E_{Tot}}$[c]	*ortho*	–105.2	–27.2	–56.9	–14.9
		(–98.9)	(–24.0)	(–52.8)	(–12.5)
	para	–102.0	–53.2	–56.0	–[b]
		(–95.8)	(–48.0)	(–51.9)	
$\mathbf{E_{Def}}$	*ortho*	12.9	74.7	7.8	1.4
	para	12.9	4.3	8.0	–[b]
		Adenine		**Guanine**	
		WC	**Hoogsteen**	**WC**	**Hoogsteen**
$\mathbf{E_{Int}}$[a]	*ortho*	–61.1	–20.4	–54.9	–56.1
	para	–60.7	–35.3	–49.6	–49.9
$\mathbf{E_{Tot}}$[c]	*ortho*	–53.4	–17.6	–52.9	1.5
		(–47.4)	(–15.0)	(–48.5)	(4.7)
	para	–53.1	–32.7	–47.7	–47.4
		(–47.1)	(–28.4)	(–43.4)	(–42.8)
$\mathbf{E_{Def}}$	*ortho*	7.7	2.9	2.0	57.6
	para	7.6	2.7	1.9	2.5

[a] The binding strength calculated as the energy difference between the complex and the monomers in the complex geometry (excluding a zero-point vibrational energy correction). [b] No viable thymine pair was identified (see Ref. 46). [c] The binding strength calculated as the energy difference between the complex and the individually optimized monomers (excluding a zero-point vibrational energy correction). Binding energies including (0.9806) scaled zero-point vibrational energy corrections are provided in parentheses.

C. Nevertheless, the quality of the adduct:C Hoogsteen pair is dependent on the structural fit within a DNA helix, which is difficult to evaluate using the nucleobase model discussed in this section. This emphasizes the importance of using a larger model for determining base-pairing preferences.

When mismatches are considered, the next strongest Hoogsteen bonding mode of *o*-PhOH-G occurs with G (E_{Int} = –56 kJ mol^{-1}, Figure 3.3b), which is the strongest pair with *syn* natural dG. However, T and A both form very weak Hoogsteen pairs with *o*-PhOH-G (–16 and –20 kJ mol^{-1}, respectively). These mismatches are both slightly destabilized relative to the corresponding natural dG pairs, and are much weaker than the natural dG:C (–113.5 kJ mol^{-1}) and dA:T (–54.1 kJ mol^{-1})[44] contacts. Furthermore, as found for natural dG, the propeller-twist and buckle distortions are significant when the adduct binds to T or G, which suggests that neither pair will properly fit into a helix. This suggests that unlike the C Hoogsteen pair, the presence of the three possible mismatches will likely destabilize DNA helices. Thus, calculations using a nucleobase model indicate that if the *o*-PhOH-G appears in the *syn* conformation, then a strong Hoogsteen complex will only be formed with C.

It is particularly interesting that *o*-PhOH-G binds strongly to C regardless of whether the *anti* or *syn* conformation is adopted. This may suggest that the adduct can exist in either conformation in a DNA helix without causing significant destabilization or mutations. However, closer examination of the geometry of the adduct monomer reveals larger structural changes when bound to C using the Hoogsteen face compared with the Watson-Crick face. Specifically, the monomer in the Watson-Crick pair resembles the global minimum of the (isolated) *o*-PhOH-G nucleobase adduct, which is planar and strongly stabilized by the presence of an O–H···N7 intramolecular hydrogen bond. However, in the Hoogsteen base pair, the hydroxyl group in the adduct rotates and causes a twisted conformation about θ due to a repulsive H–O···N7 interaction, which corresponds to a significantly higher-energy conformation of the nucleobase (by 52 kJ mol^{-1} relative to the planar global minimum, see Figure 2.6). Since the adduct geometry is significantly affected by base-pair formation, the deformation energy must be examined in order to accurately compare the relative binding strengths.

As found for natural dG (Table 3.1), the adduct geometry is not greatly affected by Watson-Crick base-pair formation. Therefore, the binding strengths decrease by only 2 – 13 kJ mol^{-1} when the deformation energy is included in the interaction energy (Table 3.2). Specifically, the (E_{Tot}) binding strength of the adduct:C pair decreases to –105 kJ mol^{-1}, while those of the mismatched pairs with T, A and G fall within a range of –53 to –57 kJ mol^{-1} (Table 3.2). This confirms that C leads to the strongest Watson-Crick pair, and the previous conclusions hold regardless of whether the deformation energy is included in the calculated Watson-Crick binding strengths.

The effect of accounting for the monomer deformation energy in the total interaction energy is much larger for the Hoogsteen pairs of *o*-PhOH-G than for the corresponding Watson-Crick pairs or the natural Hoogsteen pairs. For example, the total (E_{Tot}) binding strength of the C pair is only –27 kJ mol^{-1}, which represents a decrease in stability of 74 kJ mol^{-1} due to rotation of the hydroxyl group in the adduct upon binding (Table 3.2). The adduct:G pair leads to a similar conformational change, and a significant decrease in the binding energy. In fact, when the deformation energy is included, the adduct:G interaction is (1.5 kJ mol^{-1}) destabilizing even though the natural dG pair remains (–51 kJ mol^{-1}) stable. The T and A interactions with the adduct are only slightly affected by the deformation energy and remain weakly favorable (–15 and –18 kJ mol^{-1}, respectively). Therefore, even when the deformation of the adduct monomer is taken into account, C interactions with the Hoogsteen face are the strongest among all bases (Table 3.2). Nevertheless, the (E_{Tot}) binding strength of this pair is much weaker (by 27 kJ mol^{-1}) than a typical dA:T interaction (–54.1 kJ mol^{-1}).[44] It is also interesting to note that the deformation energy for the natural dG:C pair is less than 2 kJ mol^{-1}, and therefore the total C interaction strength (including deformation energy) for the adduct (–27 kJ mol^{-1}) is now weaker than for the natural nucleoside (–36 kJ mol^{-1}). Thus, it is anticipated that any pair formed with

the Hoogsteen face of the adduct will destabilize DNA relative to the corresponding natural sequence.

In summary, when the *o*-PhOH-G adduct monomer deformation energy is included in the calculated interaction strengths, the Hoogsteen pairs are greatly destabilized relative to the Watson-Crick pairs. It is anticipated that the strong C Watson-Crick interaction compared to the Hoogsteen interaction will stabilize the *anti* conformation of *o*-PhOH-G in DNA. In addition, even when mismatches are considered, the *anti* conformation of *o*-PhOH-G will form stronger base pairs than the *syn* conformation. This suggests that hydrogen-bonding interactions will favour adoption of the *anti* conformation in DNA helices, and if the *syn* conformation is induced by the bulky group, destabilization of the helix will result from loss of hydrogen-bond stability.

3.3.2.2 Nucleoside Model

The preceding discussion indicates that deformation of the adduct monomer upon base-pair formation has a large effect on the calculated binding strength. However, Chapter 2 shows that the monomer geometry also changes in the presence of the DNA sugar moiety. Specifically, inclusion of deoxyribose induces a twist about the θ dihedral angle by 21° in the lowest energy *anti* conformation of the isolated *ortho* nucleoside adduct, and 25° in the lowest energy *syn* structure. Since it is unknown how the twist will affect hydrogen-bond formation or vice versa, the following section considers the effects of using a nucleoside model on the calculated hydrogen-bonded structures and strengths.

The structures of the Watson-Crick (*anti* conformation) pairs involving the adduct nucleoside are very similar to the corresponding natural complexes, where the complex with C has a geometry most suitable for a typical B-DNA helix (Figure 3.4a). More importantly, *o*-PhOH-dG remains twisted about θ by ~21° in all Watson-Crick hydrogen-bonded complexes compared to θ = 180° (planar) when the nucleobase model is considered

(Figure 3.3a). Nevertheless, the (E_{Int}) interaction energies of the two models deviate by less than 2 kJ mol^{-1} (Tables 3.2 and 3.3), which indicates that the (E_{Int}) Watson-Crick binding strengths are not affected by the degree of twist. Therefore, C remains the strongest Watson-Crick bonding partner upon model extension (Table 3.3). Furthermore, the effect of including the monomer deformation energy is similar for the nucleoside and nucleobase models, as well as for all base pairs. Specifically, the total (E_{Tot}) Watson-Crick binding strength for each nucleoside base pair is smaller by only 1 – 12 kJ mol^{-1} (Table 3.2), where the most significant difference occurs for the C complex. Therefore, as found for the nucleobase model, the binding strengths calculated using the adduct nucleoside follow the same trend as those for the natural nucleoside and support preferential pairing with C with little destabilization to the helix.

The (E_{Int}) binding strengths of pairs involving the Hoogsteen face (*syn* conformation) of the *o*-PhOH-dG nucleoside adduct (Table 3.3) are smaller by up to only 3 kJ mol^{-1} than the nucleobase (Table 3.2). Furthermore, the geometries of the complexes determined using the nucleoside (Figure 3.4b) and nucleobase (Figure 3.3b) models are similar with the exception of the twist about θ. Interestingly, the degree of twist depends on the model and the base pair considered, which implies that inclusion of the deformation energy will have a varying effect on the calculated binding strengths of some nucleoside pairs compared with the nucleobase model. For example, in the C complex, the nucleoside adduct is twisted about θ by 40°, which leads to a decrease in the binding strength of 60 kJ mol^{-1} (Table 3.3) upon inclusion of monomer deformation energy, while the nucleobase adduct is twisted by 26°, which results in a 74 kJ mol^{-1} reduction in binding (Table 3.2). This difference leads to a 11 kJ mol^{-1} stronger (E_{Tot}) total interaction energy with C for the nucleoside model compared with the nucleobase model. Although the T and A complexes are not significantly affected by the model or deformation energy,

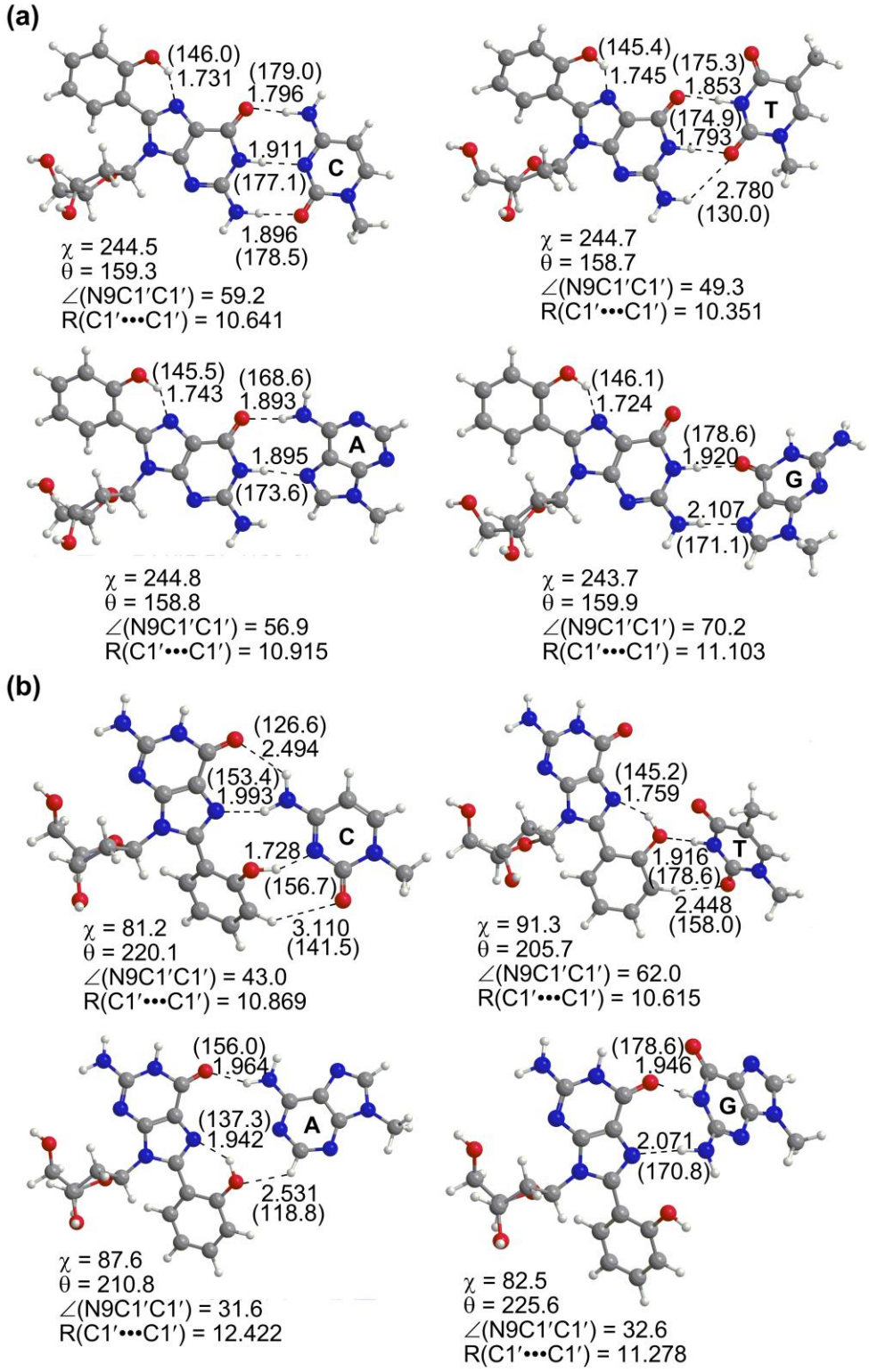


Figure 3.4 B3LYP/6-31G(d) structures (distances in Å, angles in degrees) for complexes between the (a) Watson-Crick (*anti*) and (b) Hoogsteen (*syn*) faces of the *o*-PhOH-dG (nucleoside) adduct and each of the four natural DNA nucleoside models.

Table 3.3 B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) binding strengths (kJ mol^{-1}) between the Watson-Crick (*anti*) or Hoogsteen (*syn*) face of the *o*- or *p*-PhOH-dG (nucleoside) adduct and the four natural DNA nucleoside models.

		Cytosine		Thymine	
		WC	**Hoogsteen**	**WC**	**Hoogsteen**
$\mathbf{E_{Int}}$[a]	*ortho*	–117.3	–98.9	–62.7	–16.8
	para	–113.9	–54.3	–61.8	– [b]
$\mathbf{E_{Tot}}$[c]	*ortho*	–105.4 (–99.2)	–38.6 (–35.9)	–55.8 (–52.0)	–15.3 (–12.8)
	para	–100.6 (–94.7)	–49.8 (–44.2)	–53.7 (–50.0)	– [b]
$\mathbf{E_{Def}}$	*ortho*	11.9	60.2	6.9	1.5
	para	13.3	4.5	8.1	– [b]
		Adenine		**Guanine**	
		WC	**Hoogsteen**	**WC**	**Hoogsteen**
$\mathbf{E_{Int}}$[a]	*ortho*	–62.6	–21.1	–55.5	–57.5
	para	–61.9	–34.2	–50.1	–53.7
$\mathbf{E_{Tot}}$[c]	*ortho*	–55.5 (–49.7)	–16.6 (–13.8)	–54.4 (–50.0)	–12.3 (–10.1)
	para	–53.9 (–48.3)	–31.5 (–26.6)	–48.1 (–43.7)	–51.2 (–46.4)
$\mathbf{E_{Def}}$	*ortho*	7.0	4.5	1.1	45.2
	para	8.0	2.7	2.0	2.5

[a] The binding strength calculated as the energy difference between the complex and the monomers in the complex geometry (excluding a ZPVE correction). [b] No viable thymine pair was identified (see Ref. 46). [c] The binding strength calculated as the energy difference between the complex and the individually optimized monomers (excluding a ZPVE correction). Binding energies including (0.9806) scaled ZPVE corrections are provided in parentheses.

inclusion of the deformation energy leads to a 14 kJ mol^{-1} stronger (E_{Tot}) binding strength for the adduct:G pair when the larger nucleoside model is used.

Comparison of binding strengths that neglect (E_{Int}) and include (E_{Tot}) the monomer deformation energy indicate that changes in the monomer geometry upon binding have a significant effect on the overall base-pair binding strength. This effect is in part due to the

large conformational (twist) flexibility in the adduct, which is dependent upon the presence of the deoxyribose sugar moiety. Therefore, despite the popularity of using small, nucleobase models to study hydrogen-bonding patterns,[9,19,23-26,31,34,36] future work should consider adduct models at least as large as the nucleoside in order to calculate accurate interaction energies for bulky C8 adducts which may undergo large structural changes with attachment of the deoxyribose sugar moiety. Although the present model cannot account for all environmental factors in DNA helices (solvent, flanking bases, etc.), expanding the computational model to yield structures even more similar to those found in DNA will further improve the accuracy of the calculated deformation energy. However, since the sugar moiety has a large effect on θ, it is anticipated that the nucleoside model provides a better estimate of the deformation energy than the nucleobase model, and that deformation energy will continue to be less important as the model size increases.

The Hoogsteen mismatches involving the *o*-PhOH-dG are significantly destabilized relative to the corresponding pairs with the natural nucleoside upon inclusion of the monomer deformation energy. In fact, the (E_{Tot}) binding strengths with A, G or T are diminished to the point where these pairs are not stable enough to be feasible in a DNA helix (Table 3.3). On the other hand, the adduct:C complex is actually slightly (2.6 kJ mol^{-1}) more stable than the natural dG:C Hoogsteen pair. Furthermore, the R(C1′···C1′) distance and the ∠(N9C1′C1′) angle in the adduct complex are closer to the values expected for DNA than the natural nucleoside pair. Therefore, despite the similarity in stability of the Hoogsteen C pair involving *o*-PhOH-dG and dG, the shape of the adduct complex is typical of a natural Watson-Crick pair, and the *syn* conformation of the adduct may consequently be less destabilizing than that of the natural nucleoside when paired with C in DNA helices. This example illustrates that the presence of the bulky group can improve the Hoogsteen binding, and reemphasizes the importance of considering both the binding strength and

structure when evaluating base pairs, as well as the use of larger models that afford more detailed structural information.

In summary, as found for the nucleobase model, deformation of the *o*-PhOH-dG adduct upon base-pair formation has a large effect on the calculated binding strengths, where all base pairs are predicted to be less stable when the deformation energy is included. In addition to the valuable structural information obtained from the larger model, the significant differences in the deformation energy between the two models provides further support for using at least a nucleoside model in future studies of other flexible, bulky adducts. Given that the Watson-Crick interaction is significantly stronger (by 67 kJ mol^{-1}), the *anti* conformation of the damaged base may be stabilized in the helix, as reported for the AF-dG adduct when paired opposite C.[2] Nevertheless, the *syn* conformation also forms a reasonably strong (–38.6 kJ mol^{-1}), and structurally viable, base pair with C, and may still afford a stable double helix should other factors (stacking interactions, major groove contacts) induce the *syn* conformation. Most importantly, C is the base favored by *o*-PhOH-dG regardless of the orientation about the glycosidic bond. Therefore, if the small model implemented in this chapter accurately predicts B-DNA bonding preferences, the results suggest that misincorporation is not expected to occur for *o*-PhOH-dG.

3.3.3 *p*-PhOH-dG Adduct

Since the above section concluded that the nucleoside model is required to accurately calculate the binding strengths and provide detailed structural information of the *o*PhOH-dG adduct pairs, only the results obtained with the nucleoside model for the *p*-PhOH-dG adduct will be discussed in this section. In addition, only the total (E_{Tot}) binding strengths for the *p*-PhOH-dG complexes that include the monomer deformation energy will be discussed since these values for *o*-PhOH-dG were concluded to be the most accurate for this model. However, the interaction energies without the deformation energy included for

the nucleoside model are also available in Table 3.3. Additionally, the binding strengths and geometries calculated with the nucleobase model can be found in Table 3.2 and Figure 3.5, respectively, which show that the smaller model yields binding strengths and geometries similar to the nucleoside model for the *p*-PhOH-dG adduct.[45]

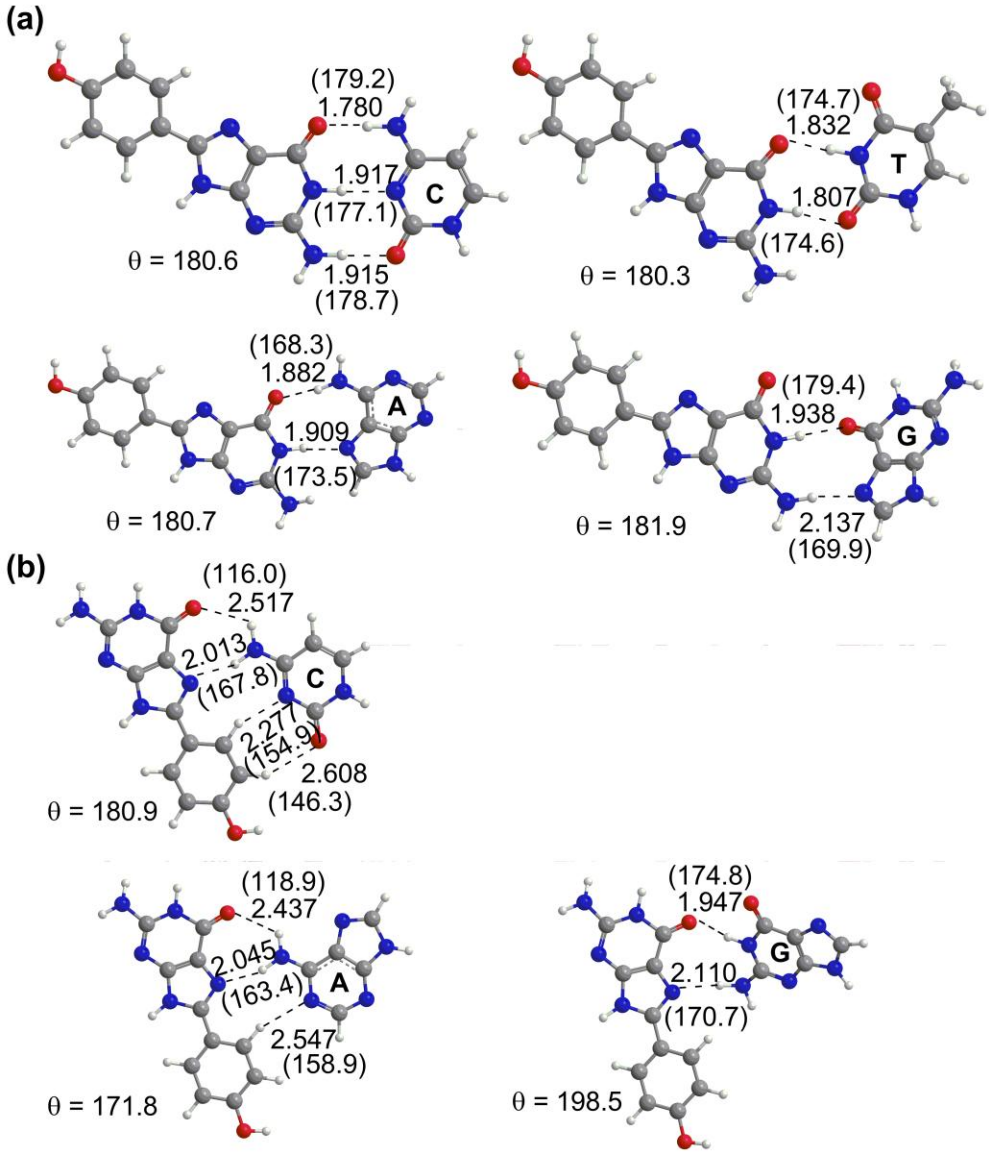


Figure 3.5 B3LYP/6-31G(d) structures (distances in Å, angles in degrees) for complexes between the (a) Watson-Crick (*anti*) and (b) Hoogsteen (*syn*) faces of the *p*-PhOH-G (nucleobase) adduct and each of the four natural DNA nucleobases.

As discussed in Chapter 2, *p*-PhOH-dG lacks the stabilizing O–H···N7 intramolecular hydrogen bond contained in *o*-PhOH-dG (Figure 3.1), which leads to a larger twist in the isolated nucleoside with respect to θ in both the *anti* and *syn* conformations (θ = 31 or 34° for *p*-PhOH-dG compared to 21 or 25° for *o*-PhOH-dG *anti* or *syn* conformations, respectively). As discussed for *o*-PhOH-dG, deviations in the twist angle can change the geometries of the hydrogen-bonding pairs. It is not intuitive how differences in the twist in the *o*- and *p*-PhOH-dG adduct monomers, as well as the different location of the hydroxyl group, will affect the overall base-pairing preference of the damaged base.

The *anti* conformation of *p*-PhOH-dG is expected to form hydrogen-bonded pairs with the natural nucleosides in an analogous manner to *o*-PhOH-dG since the Watson-Crick face is unaffected by the damage. Indeed, the two adducts form structurally similar hydrogen-bonding interactions with each natural nucleoside model (Figures 3.4a and 3.6a) and therefore the calculated binding strengths are within 6 kJ mol^{-1} (Table 3.3). The largest deviation occurs between the G pairs due to greater propeller-twist and buckle distortions in the *p*-PhOH-dG adduct complex. All structures and binding strengths are also close to those for the corresponding natural dG pairs. Most importantly, as found for natural dG and *o*-PhOH-dG, *p*-PhOH-dG interacts most strongly with C, which also forms the pair with the most appropriate structure for a DNA helix.

Unlike *o*-PhOH-dG, the Hoogsteen-bonding face of *syn p*-PhOH-dG (Figure 3.1c) has no additional hydroxyl hydrogen-bonding sites compared with natural G. Instead, the bulky phenoxyl group forms additional weak C–H···X interactions upon binding to the natural bases. These interactions actually preclude the formation of a viable T pair.[46] Furthermore, *p*-PhOH-dG has slightly larger Hoogsteen binding strengths with the other natural bases than dG (by up to 13.8 kJ mol^{-1}). The largest energetic difference occurs for the C pair, which also has a significant (R(C1′···C1′), ∠(N9C1′C1′)) structural improvement (10.5 Å, 45°) over

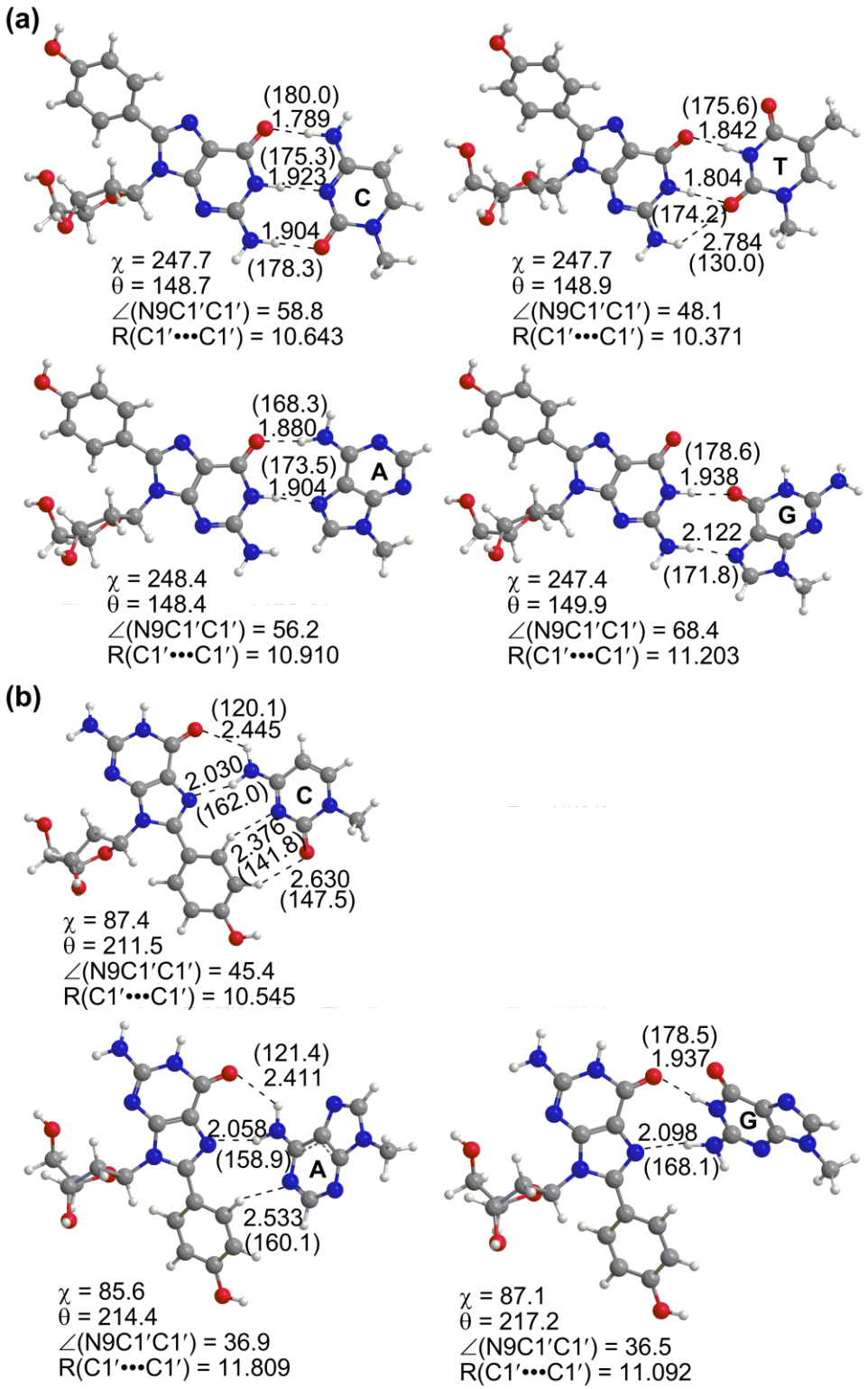


Figure 3.6 B3LYP/6-31G(d) structures (distances in Å, angles in degrees) for complexes between the (a) Watson-Crick (*anti*) and (b) Hoogsteen (*syn*) faces of the *p*-PhOH-dG (nucleoside) adduct and each of the four natural DNA nucleoside models.

the (*syn*) natural dG complex (11.7 Å, 30°). Although the adduct:A pair is distorted, structural improvement is also seen in this complex compared to the dG mismatch. Therefore, the presence of the bulky group improves the quality of these Hoogsteen complexes relative to the corresponding pair composed of natural bases and therefore increases the likelihood of mismatch incorporation compared to natural dG. As found for natural dG, the guanine Hoogsteen complex has the largest binding strength despite significant propeller-twist and buckle distortions (Figure 3.6b). However, for *p*-PhOH-dG, the C complex is similar in stability and is less distorted than the G complex. Since calculations on other non-natural base pairs indicate that calculated binding strengths may not be significantly affected even if environmental constraints enforce a less distorted complex,[47] the results indicate that both C and G may form viable Hoogsteen pairs of similar stability (Table 3.3). Therefore, if *p*-PhOH-dG adopts the *syn* conformation, G mismatches may be favored.

Interestingly, all Hoogsteen pairs are more stable than the corresponding *o*-PhOH-dG complexes (by 11 – 39 kJ mol^{-1}). Thus, while the Hoogsteen complexes of *o*-PhOH-dG are quite unstable relative to the Watson-Crick pairs, there is a smaller difference in stability between the *p*-PhOH-dG Hoogsteen and Watson-Crick complexes (Table 3.3). These results suggest a preference for all nucleobases to form Watson-Crick pairs with *o*-PhOH-dG, but cannot indisputably predict the preference for *p*-PhOH-dG. Indeed, if other factors (sterics, stacking, major groove contacts) cause the PhOH adducts to adopt the *syn* conformation in the DNA helix, *p*-PhOH-dG may be less destabilizing than *o*-PhOH-dG.

In summary, among all natural bases, C forms the most structurally feasible pairs with both the Watson-Crick and Hoogsteen faces of *p*-PhOH-dG, where the Watson-Crick complex is more stable. However, both C and G form relatively stable Hoogsteen pairs that are comparable in strength to the natural A:T complex. Furthermore, although Watson-

Crick binding is strongly favored for natural dG and *o*-PhOH-dG, the difference between the Watson-Crick and Hoogsteen binding faces is significantly smaller for *p*-PhOH-dG. Indeed, if the phenoxyl adducts adopt the *syn* conformation in the DNA helix, *p*-PhOH-dG may be less destabilizing than the natural nucleoside or *o*-PhOH-dG. Therefore, the *anti*/*syn* preference for *p*-PhOH-dG is less clear than for *o*-PhOH-dG, and *p*-PhOH-dG may readily form base pairs in either conformation with more than one natural base. Larger models are needed to unambiguously determine the conformational and base-pairing preference of *p*-PhOH-dG, which will better reveal how this adduct could lead to mutations.

3.4 Conclusion

The hydrogen-bonding preferences of the Watson-Crick (*anti*) and Hoogsteen (*syn*) faces of the *o*- and *p*-PhOH-dG adducts for the four natural bases were calculated to determine the possible effect of hydrogen-bond formation on the *anti*/*syn* preference of the damaged bases, as well as predict the potential for misincorporation events upon replication. The magnitudes of the binding strengths, as well as the structural features of the complexes (R(C1′···C1′) distance, ∠(N9C1′C1′) angle, propeller-twist and buckle distortions) were evaluated and compared to those expected for natural DNA helices. Comparison of data from nucleoside and nucleobase models of the *o*-PhOH-dG adduct suggests that the nucleobase model overestimates the magnitude of the deformation energy. Therefore, even though both models have the same base-pairing trend when deformation energy is included, nucleoside models are the smallest computational models that should be used for accurate prediction of the binding strengths of flexible, bulky adducts that may undergo significant geometric changes upon complex formation.

Using the most accurate (nucleoside) models, it was predicted that *o*-PhOH-dG prefers to bind to C regardless of the hydrogen-bonding face considered, if these models accurately predict the binding arrangement adopted in DNA helices. This suggests the

o-PhOH-dG adduct may have a low potential to result in misincorporation. Although a preference for the *syn* conformation was found in Chapter 2, the present work indicates the Watson-Crick face forms significantly more stable complexes than the Hoogsteen face, and the *anti* conformation of the adduct may be favoured by hydrogen-bonding interactions in DNA helices. However, the Hoogsteen face also forms a reasonably strong and structurally feasible pair with C. Therefore, the current model suggests *o*-PhOH-dG will not readily lead to base-substitution mutations even if the *syn* conformation is adopted in DNA helices.

While *o*-PhOH-dG and the natural dG nucleoside form much stronger complexes with the Watson-Crick than the Hoogsteen face, the difference in stability is less significant for *p*-PhOH-dG. This suggests that hydrogen-bonding interactions may not prevent *p*-PhOH-dG from readily adopting multiple conformations in DNA helices. Furthermore, although the Watson-Crick face forms the most stable complex with C, the Hoogsteen face forms hydrogen bonds to both C and G that are comparable in stability to a Watson-Crick A:T pair. Therefore, this model predicts the *syn* conformation of *p*-PhOH-dG may be adopted in DNA helices if other factors also favour this conformation. In this instance, this damaged base is able to bind to multiple partners. This suggests a possible mechanism by which *p*-PhOH-dG may lead to mutations. Nevertheless, the isolated *para* adduct:G complex is significantly distorted and the effects of other environmental considerations in the DNA strand (such as stacking, sterics, major groove contacts) on the binding strength must be investigated to conclusively determine the conformational and binding preferences of this adduct. Therefore, future chapters explore the conformational preferences of both adducts in greater depth using models of gradually increasing size. Specifically, a nucleotide model will be considered in Chapter 4, while a dinucleoside monophosphate model is presented in Chapter 5. Finally, conformational and base-pairing preferences will be simultaneously investigated using a duplex model in Chapter 6, where the base pairs predicted to form the

strongest binding interactions with the nucleoside model will be incorporated in the larger model.

3.5 References

(1) Wang, F.; Elmquist, C. E.; Stover, J. S.; Rizzo, C. J.; Stone, M. P. *Biochemistry* **2007**, *46*, 8498-8516.

(2) Liang, F. T.; Cho, B. P. *Chem. Res. Toxicol.* **2011**, *24*, 597-605.

(3) Gubala, V.; Betancourt, J. E.; Rivera, J. M. *Org. Lett.* **2004**, *6*, 4735-4738.

(4) Sessler, J. L.; Jayawickramarajah, J.; Sherman, C. L.; Brodbelt, J. S. *J. Am. Chem. Soc.* **2004**, *126*, 11460–11461.

(5) Hernandez, B.; Soliva, R.; Luque, F. J.; Orozco, M. *Nucleic Acids Res.* **2000**, *28*, 4873-4883.

(6) Hu, X. B.; Li, H. R.; Ding, J. Y.; Han, S. J. *Biochemistry* **2004**, *43*, 6361-6369.

(7) Volk, D. E.; Thiviyanathan, V.; Somasunderam, A.; Gorenstein, D. G. *Org. Biomol. Chem.* **2007**, *5*, 1554-1558.

(8) Sponer, J.; Sabat, M.; Gorb, L.; Leszczynski, J.; Lippert, B.; Hobza, P. *J. Phys. Chem. B* **2000**, *104*, 7535-7544.

(9) Sponer, J.; Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142-10151.

(10) Fonseca Guerra, C.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *J. Am. Chem. Soc.* **2000**, *122*, 4117-4128.

(11) Giudice, E.; Varnai, P.; Lavery, R. *Nucleic Acids Res.* **2003**, *31*, 1434-1443.

(12) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.

(13) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502-16513.

(14) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.;

Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03, Revisions C.02 and D.01; Gaussian, Inc.: Wallingford, CT, 2004.

(15) Deepa, P.; Kolandaivel, P. *J. Biomol. Struct. Dyn.* **2008**, *25*, 733-746.

(16) Gould, I. R.; Kollman, P. A. *J. Am. Chem. Soc.* **1994**, *116*, 2493-2499.

(17) Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608-15613.

(18) Kabelac, M.; Hobza, P. *J. Phys. Chem. B* **2001**, *105*, 5804-5817.

(19) Kryachko, E. S.; Sabin, J. R. *Int. J. Quantum Chem.* **2003**, *91*, 695-710.

(20) Liu, H.; Gauld, J. W. *Phys. Chem. Chem. Phys.* **2009**, *11*, 278-287.

(21) Mathews, D. H.; Case, D. A. *J. Mol. Biol.* **2006**, *357*, 1683-1693.

(22) Mo, Y. R. *J. Mol. Model.* **2006**, *12*, 665-672.

(23) Mohajeri, A.; Nobandegani, F. F. *J. Phys. Chem. A* **2008**, *112*, 281-295.

(24) Noguera, M.; Sodupe, M.; Bertran, J. *Theor. Chem. Acc.* **2004**, *112*, 318-326.

(25) Padermshoke, A.; Katsumoto, Y.; Masaki, R.; Aida, M. *Chem. Phys. Lett.* **2008**, *457*, 232-236.

(26) Quinn, J. R.; Zimmerman, S. C.; Del Bene, J. E.; Shavitt, I. *J. Am. Chem. Soc.* **2007**, *129*, 934-941.

(27) Radisic, D.; Bowen, K. H.; Dabkowska, I.; Storoniak, P.; Rak, J.; Gutowski, M. *J. Am. Chem. Soc.* **2005**, *127*, 6443-6450.

(28) Sharma, P.; Mitra, A.; Sharma, S.; Singh, H.; Bhattacharyya, D. *J. Biomol. Struct. Dyn.* **2008**, *25*, 709-732.

(29) Bandyopadhyay, D.; Bhattacharyya, D. *Biopolymers* **2006**, *83*, 313-325.

(30) Kelly, R. E. A.; Kantorovich, L. N. *J. Phys. Chem. C* **2007**, *111*, 3883-3892.

(31) Monajjemi, M.; Chahkandi, B. *J. Mol. Struct. THEOCHEM* **2005**, *714*, 43-60.

(32) Pacureanu, L.; Simon, Z. *Int. J. Quantum Chem.* **2010**, *110*, 1295-1305.

(33) Priyakumar, U. D.; MacKerell, A. D. *Chem. Rev.* **2005**, *106*, 489-505.

(34) Sharma, P.; Mitra, A.; Sharma, S.; Singh, H. *J. Chem. Sci.* **2007**, *119*, 525-531.

(35) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Mol. Struct. THEOCHEM* **2001**, *573*, 43-53.

(36) Sponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers* **2001**, *61*, 3-31.

(37) , P.; Hobza, P. *J. Phys. Chem. A* **2005**, *109*, 1131-1136.

(38) Asensio, A.; Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A* **2003**, *107*, 6441-6443.

(39) Kelly, R. E. A.; Lee, Y. J.; Kantorovich, L. N. *J. Phys. Chem. B* **2005**, *109*, 22045-22052.

(40) Robertazzi, A.; Platts, J. A. *J. Phys. Chem. A* **2006**, *110*, 3992-4000.

(41) Roy, A.; Panigrahi, S.; Bhattacharyya, M.; Bhattacharyya, D. *J. Phys. Chem. B* **2008**, *112*, 3786-3796.

(42) Johnson, R. E.; Haracska, L.; Prakash, L.; Prakash, S. *Mol. Cell. Biol.* **2006**, *26*, 6435-6441.

(43) Benitez, B. A. S.; Arora, K.; Balistreri, L.; Schlick, T. *J. Mol. Biol.* **2008**, *384*, 1086-1097.

(44) Calculated at the B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) level.

(45) The model size affects the deformation energy for the *ortho*, but not *para*, adduct since the favourable O–H•••N7 hydrogen bond in the *ortho* adduct monomer is disrupted upon base pairing and this disruption is greater for the nucleobase than the nucleoside model. In contrast, there is no similar disruption in the most favourable geometry of the *para* adduct upon base pairing.

(46) This is due to steric clashes between the phenoxyl group and N3–H of thymine (i.e.; C–H•••H–N3 contacts). Indeed, no Hoogsteen *para* adduct–thymine pair is reported in the present study because optimizations result in highly distorted, unrealistic interactions between the *para* hydroxyl group and O4 of thymine. In contrast, the Hoogsteen natural guanine–thymine pair forms a C8–H•••O2 hydrogen bond, which occurs due to distortion between the bases to minimize repulsion between O6 (G) and O4 (T).

(47) Rutledge, L. R.; Wheaton, C. A.; Wetmore, S. D. *Phys. Chem. Chem. Phys.* **2007**, *9*, 497-509.

4 Chapter 4: Conformational Flexibility of C-Linked PhOH-dG Adducts using a Nucleotide Model[a]

4.1 Introduction

Chapter 2 concluded that the *o*- and *p*-PhOH-dG adducts prefer a *syn* conformation due to the presence of a strong O5′–H···N3 hydrogen bond, as well as the destabilization of the *anti* structure due to positioning of the bulky group directly above the sugar moiety. However, it is not clear if the preference of the nucleoside to adopt the *syn* structure applies to the DNA duplex, where the stabilizing O5′–H···N3 hydrogen-bond contact is no longer possible due to the presence of a phosphate group at C5′. Indeed, previous literature suggests that the preference for the *anti* versus *syn* conformation of the natural nucleosides is dependent on the presence of intramolecular hydrogen bonds, as well as the computational model employed.[1-5] Therefore, it may not be possible to extrapolate conclusions drawn for the nucleoside model to the physiological system. This is significant since experiments performed to date have only used the nucleoside and nucleobase adducts.[6-8] Therefore, more careful consideration of the *anti*/*syn* relative energies is required, which will clarify the conformational preferences in biologically-relevant systems and permit accurate assessment of the mutagenic potential of the phenoxyl adducts.

In this chapter, the *anti*/*syn* preference of the *o*- and *p*-PhOH-dG adducts will be investigated by considering larger model systems that better mimic the environment in DNA helices. Specifically, the relative energies of *anti*/*syn* conformations using a nucleoside model is first adjusted by including a geometric constraint, which prevents interactions between the base and the sugar moiety that cannot occur in the helix. Subsequently, the nucleoside model will be expanded to a nucleotide model by including the 5′-monophosphate group, which will also provide insight into the effect of potential

[a] Reprinted in part with permission from Millen, A. L.; Manderville, R. A.; Wetmore, S. D. *J. Phys. Chem. B* **2010**, *114*, 4373-4382. Copyright 2010 American Chemical Society.

interactions between the base and the DNA phosphate backbone on the conformational preference.

Only one study in the literature to date calculated the *anti*/*syn* preference of a bulky adduct using a nucleotide model.[9] Furthermore, computational modeling of the *anti*/*syn* preference of even the natural nucleotides has proven challenging,[10-14] where the correct *anti*/*syn* energy trend for natural dG 5′-monophosphate (dGMP) has only been found using very specific computational models.[15] Since these models assume prior knowledge of the biologically-relevant nucleotide conformation, which is unavailable for most damaged systems, these methods cannot be applied to the *o*- and *p*-PhOH-dG adducts. Therefore, before studying the *anti*/*syn* energy difference for the *o*- and *p*-PhOH-dG adducts, a suitable computational approach will be first be identified by investigating natural dGMP. The methodology will then be applied to the study of the *o*- and *p*-PhOH-dGMP adducts to provide new information about the structure of *o*- and *p*-PhOH-dGMP adducts in B-DNA. Studies of the nucleoside and nucleotide adducts will aid in establishing a protocol for using small models to predict the preferred conformation of bulky damaged bases in physiological environments.

4.2 Computational Details

4.2.1 Nucleoside Model

To provide detailed information about the conformations available to the *o*- and *p*-PhOH-dG nucleoside adducts (R = H, Figure 4.1), potential energy surfaces were generated in the gas phase at the B3LYP/6-31G(d) level of theory as described in Chapter 2. Specifically, potential energy surface scans were previously performed, where the system was optimized with constrained C2′–*endo* sugar puckering, and χ (∠(O4′C1′N9C4)) and θ (∠(N9C8C10C11), Figure 2.1) angles constrained in 10° increments from 0 to 360°, while imposing an additional constraint that fixes the ∠(C4′C5′OH) dihedral angle, which

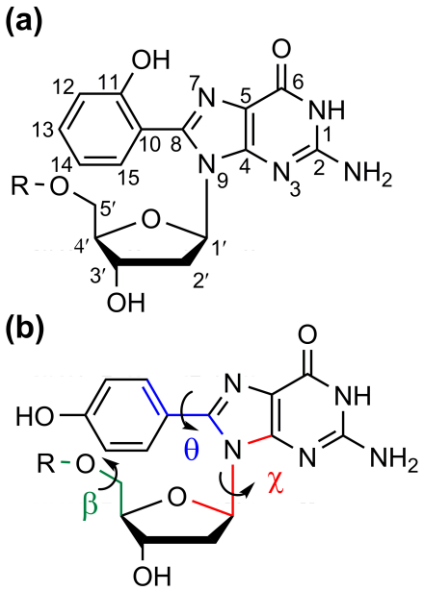


Figure 4.1The structure of the nucleoside (R = H) and nucleotide (R = PO_3^-, HPO_3, or Na^+ PO_3^-) models. The dihedral angle β (∠(C4′C5′OR)) defines the orientation of the phosphate backbone in DNA.

corresponds to the β (∠(C4′C5′OP)) angle in DNA (Figure 4.1), at 180°.[16] This additional constraint, which prevents interactions between the base and the (C5′–OH) sugar that would not occur in natural (B-DNA) oligonucleotides, has previously been used in the literature to study natural nucleosides.[2,4,5,17-19] The scans from Chapter 2 will be referred to as the β-unconstrained surfaces, while the scans generated in the current chapter will be denoted as the β-constrained surfaces. Minimum energy conformations identified on the potential energy surfaces were subsequently optimized with no constraints at the B3LYP/ 6-31G(d) level. Accurate relative energies of these fully-optimized structures were determined using B3LYP/6-311+G(2df,p) single-point calculations and include scaled (0.9806) zero-point vibrational energy[20] (ZPVE) corrections.

4.2.2 Nucleotide Model

The nucleotide model was generated by adding the 5′-monophosphate group to the lowest energy *anti* and *syn* conformations identified using the (β-constrained) nucleoside

model. As discussed in detail in the Results and Discussion, a variety of (neutral and charged) models were used to represent the phosphate group in DNA (R = PO_3^-, HPO_3 or Na^+ PO_3^-, Figure 4.1). The natural dG nucleotide was also considered in order to ensure that biologically-relevant structures were obtained with various computational models. The natural nucleotide and *o*-PhOH-dG adduct models were optimized with B3LYP and both the 6-31G(d) and 6-31+G(d,p) basis sets because of the potential importance of including diffuse functions when considering models containing phosphate anions. Due to minimal differences between the geometries obtained for *o*-PhOH-dG with the two basis sets (see Results and Discussion), *p*-PhOH-dG was only considered at the B3LYP/6-31G(d) level of theory. All models were optimized in both the gas phase and in a water environment using the PCM continuum solvation model (ε = 78.4). The reported relative energies, which include (0.9806) scaled ZPVE corrections,[20] were obtained from B3LYP/6-311+G(2df,p) single-point calculations performed in the same (gas-phase or water) environment as the corresponding optimization.

All calculations were performed using Gaussian 03.[21]

4.3 Results and Discussion

4.3.1 Nucleoside Model

The structures of C-linked PhOH-dG nucleosides were discussed in detail in Chapter 2. β-unconstrained conformational scans (Figure 2.3) reveal that *syn* minima (χ ~ 50 – 70°) of the *o*- and *p*-PhOH-dG nucleoside adducts are favored over *anti* minima (χ ~230°) by 25 – 35 kJ mol^{-1}. Furthermore, the *syn* structures may potentially fit better into a B-DNA helix. In particular, the *syn* conformations have the correct (C2′–*endo*) sugar puckering, while the *anti* conformers prefer O4′–*endo* puckering due to interactions between the bulky group and the sugar moiety. Additionally, the greater number of *syn* local minima with respect to

the θ dihedral angle suggests that the bulky group may twist with respect to the nucleobase as required to accommodate the helical environment.

Figure 4.2 displays the β-constrained potential energy surfaces of the PhOH-dG adducts. The *syn* regions of the contour plots for *o*- and *p*-PhOH-dG are similar to the surfaces in Chapter 2, where four *syn* minima can be identified for each adduct. For *o*-PhOH-dG, the differences in the *anti* regions are also minimal. However, the change in the *anti* region is more significant for *p*-PhOH-dG. Specifically, four *anti* minima are found on the β-constrained surface (Figure 4.2, right), but only two are found on the β-unconstrained surface (Figure 2.3, right). The two additional minima on the β-constrained surface correspond to rotation of the bulky phenoxyl group about θ, which is impeded in the (β-unconstrained) nucleoside model when the C5′–OH group is directed toward the natural nucleobase. This indicates that *p*-PhOH-dG may have greater flexibility in the *anti* conformation than initially predicted by the unconstrained nucleoside model, and therefore this conformation may be as readily accommodated in DNA helices as the *syn* conformer.

The most important difference between the potential energy surfaces (PES) with and without the constraint on the C5′–OH group is the relative energy of the *syn* and *anti* minima. Upon inclusion of the β constraint, the *anti* region becomes more energetically accessible for both adducts. For example, while the four *syn* minima of *o*-PhOH-dG are all lower in energy than the two *anti* minima on the β-unconstrained surface (Figure 2.3, left), the *anti* minima are lower in energy than two of the *syn* minima on the β-constrained surface (Figure 4.2, left). Indeed, the *anti*/*syn* energy difference is 26 kJ mol^{-1} (25 kJ mol^{-1} for *para*) on the unconstrained surface, but only 7 kJ mol^{-1} when the β-constraint is imposed. This indicates that the conformational energy trend is model dependent, and stresses the importance of considering models that better mimic the DNA environment when determining the conformational preference of damaged bases.

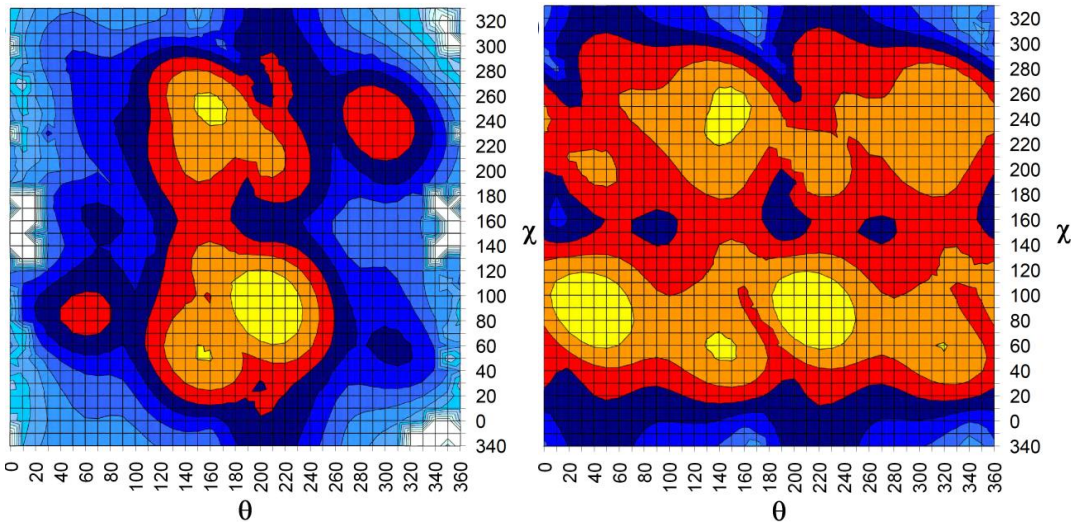


Figure 4.2 B3LYP/6-31G(d) PES for the β-constrained *o*-PhOH-dG (left) and *p*-PhOH-dG (right) adduct. The relative energy is represented by color, where yellow represents the lowest energy regions and each change in color represents an increase of 10 kJ mol^{-1}.

To further understand the *anti*/*syn* conformational preference, optimizations were performed with all constraints released for each minimum taken from the potential energy surfaces in Figure 4.2, where the resulting lowest energy *anti* and *syn* conformations for each adduct are displayed in Figure 4.3.[22] In general, for both the *anti* and *syn* structures, χ and θ values deviate from the constrained geometries by less than 5°, while β typically deviates by less than 17°. The sugar puckering of the *syn* adducts usually remains in the (C2′–*endo*) form present in B-DNA,[23] while the lowest energy *anti* structures retain O4′–*endo* puckering. The C5′–OH group remains directed away from the nucleobase (Figure 4.3) upon full optimization of the lowest energy conformers from the β-constrained surfaces.[24]

The change in the relative energies of the *anti* and *syn* conformations upon geometry relaxation is more significant than any structural change. Specifically, the *anti* structures are all stabilized upon full optimization, which results in the lowest energy *anti* minima being almost energetically equivalent to the *syn* minima, falling 0.1 – 0.2 kJ mol^{-1} below the corresponding lowest energy *syn* structure. However, due to the small *anti*/*syn*

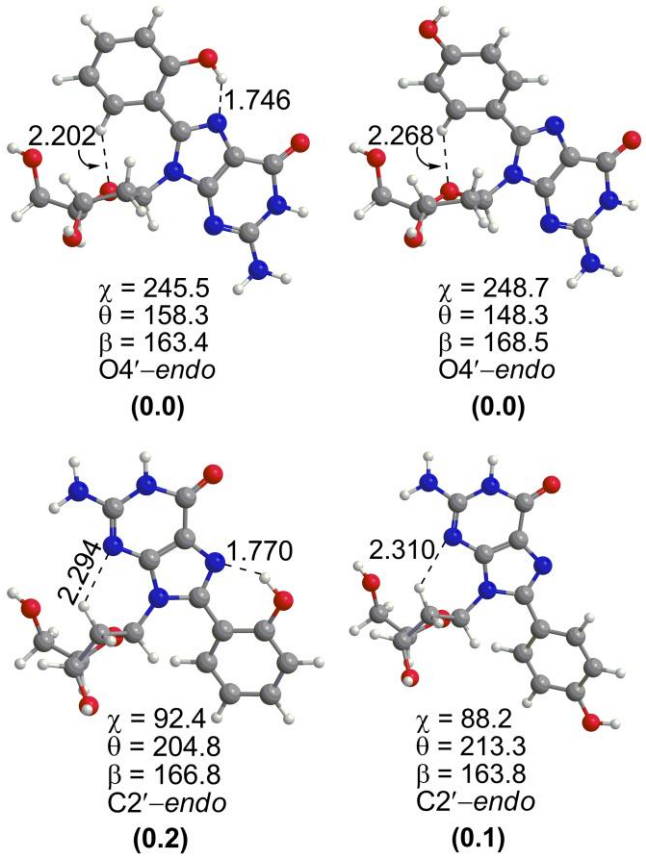


Figure 4.3 Select B3LYP/6-31G(d) hydrogen-bond lengths (Å), dihedral angles (χ, θ, and β, deg.), and sugar puckering in the minimum energy (fully-optimized) *anti* (top) and *syn* (bottom) structures obtained from the potential energy surfaces of the β-constrained *o*-PhOH-dG (right) and *p*-PhOH-dG (left) adduct. (The *anti*/*syn* relative energies (kJ mol^{-1}) from B3LYP/6-311+G(2df,p) single-point calculations are provided in bold in parentheses.)

energy difference and potential interactions between the base and the DNA phosphate group, it is still not clear from the nucleoside model whether an *anti* or a *syn* structure will predominate in a DNA duplex. Therefore, the lowest energy *syn* and *anti* conformations are further studied in the next section using a nucleotide model containing a 5′-monophosphate group.

4.3.2 Nucleotide Model

Despite drastic improvements in computational resources over the past decade, nucleotide models[5,10-12,15,19,25-32] have not been studied in the literature to the same extent as nucleoside models.[1,3,33-36] This is initially surprising since inclusion of the 5'-monophosphate group could clarify the conformational preference of DNA bases without drastically increasing the model size. However, the physiologically relevant *anti*/*syn* energy trends and geometries of nucleotides isolated by crystal structures[37] and NMR[17,38-40] have proven difficult to reproduce with computational methods. Specifically, similar to the nucleoside model, the phosphate group in nucleotide models can form interactions with nucleobases that are non-native to DNA.[5] Therefore, most computational studies that have reproduced the experimentally determined *anti*/*syn* preference for (natural) dGMP, in particular, have imposed geometric constraints on the β angle,[5] where this angle is generally fixed to 180° to mimic natural B-DNA helices.[16] Although the use of constraints derived from structural databases is a successful approach for the canonical nucleotides, no geometrical parameters for the damaged nucleotides are known. Therefore, it is not surprising that even less has been done to study the structure of nucleotides modified by a bulky substituent. Only one recent paper has studied the glycosidic-bond orientation in a bulky nucleotide adduct, which was an N7-substituted deoxyguanosine-like nucleotide.[9]

Due to the absence of experimental data for comparison, a model for the natural dG nucleotide that correctly predicts the *anti*/*syn* preference without constraints must be identified before the C-linked PhOH-dG adducts can be considered. Previous literature has used different approaches to model the structure of natural nucleotides without artificial constraints. Some groups have used deoxydinucleoside monophosphate models to successfully reproduce experimental structural data, since the phosphate is anchored in position through attachment of the second sugar moiety.[14,18,26] However, these models raise

questions regarding the sequence dependence of the conformation.[14,18,40] Although a B-DNA-like *syn* structure of dGMP can be found in gas-phase studies that use large basis sets and/or high levels of theory,[27,30] an *anti* energetic preference is not predicted.[27] Furthermore, the computational methods employed are not suitable for studying larger (damaged) nucleotides. The first computational study to correctly predict the *anti*/*syn* preference of the natural dGMP without the use of constraints used a cationic model, where a magnesium ion complexed to four explicit water molecules is coordinated to N7 of guanine and thereby anchors the nucleotide in the *anti* conformation.[15] Although this model is valid for interpreting some enzyme active sites, it is not relevant for understanding the structures of DNA duplexes.

In addition to questions regarding the implementation of geometrical constraints, there are many different ways to model the DNA phosphate moiety, which can be represented by a variety of overall neutral and charged states. When the phosphate is modeled as a neutral (H_2PO_4) group, the preferred conformation is predicted to be the *syn* conformation for all four natural nucleotides.[11,12] Monoanionic phosphate models (HPO_4^-) accurately predict dAMP, dCMP, and dTMP to prefer the *anti* conformation, but consistently predict the *syn* conformation for dGMP due to the presence of strong intramolecular hydrogen bonds between the base and the phosphate group.[10,30,31] Even though this may be the preferred gas-phase conformation for this model, these results are not relevant to biological systems. Models that include sodium counterions to neutralize the phosphate group (Na^+ HPO_4^-) give a slightly better reproduction of crystal structure data due to a net zero charge in the crystals.[26,29] Although dianionic phosphate models only describe *anti* minima for all four canonical nucleotides due to repulsion between the base and the phosphate group, the high net charge in the system makes these models unrealistic representations of DNA.[12] Therefore, in the current chapter, dGMP is modeled using the

three most commonly accepted representations of the DNA phosphate backbone: a neutral (H_2PO_4) model, a neutral sodium counterion (Na^+ HPO_4^-) model, and an anionic (HPO_4^-) model.

4.3.2.1 Gas-Phase Structure of the Natural dG Nucleotide

Although the discussion below focuses on the structure of natural dGMP obtained with the 6-31G(d) basis set, the same conclusions hold true for geometries obtained with 6-31+G(d,p), where the corresponding geometrical parameters are provided in parentheses in Figure 4.4. In all gas-phase structures (Figure 4.4), the additional proton on the phosphate group in the neutral model is not involved in interactions with the base or sugar moiety, and the only sodium interactions in the counterion model involve equivalent contacts with the two unprotonated oxygen atoms in the phosphate group (R(Na···O) ~ 2.2 – 2.4 Å).

The three (gas-phase) *anti* structures (Figure 4.4, top) have χ angles between 250 – 265°, which are within the range of values expected for B-DNA.[4] All three models form hydrogen bonds between C8–H in the base and a phosphate oxygen, as well as weaker interactions between C2′–H and the same phosphate oxygen. Both of these interactions have been previously reported in the literature.[11,27] Similar interactions between the nucleobase and phosphate backbone occur in all models since χ is similar for the counterion and neutral models, and differs by only 13 – 16° for the anionic model. However, although the neutral and counterion models form similar hydrogen bonds between the phosphate and the sugar moiety, the anionic model differs considerably. Specifically, the β angle is rotated to 75 – 88° in the neutral and counterion models, which results in a non-native hydrogen bond between the terminal protonated phosphate oxygen and O4′, while β is rotated in the opposite direction to 261° in the anionic model, which leads to interactions between the unprotonated phosphate oxygen and C3′–H in the sugar. These phosphate-

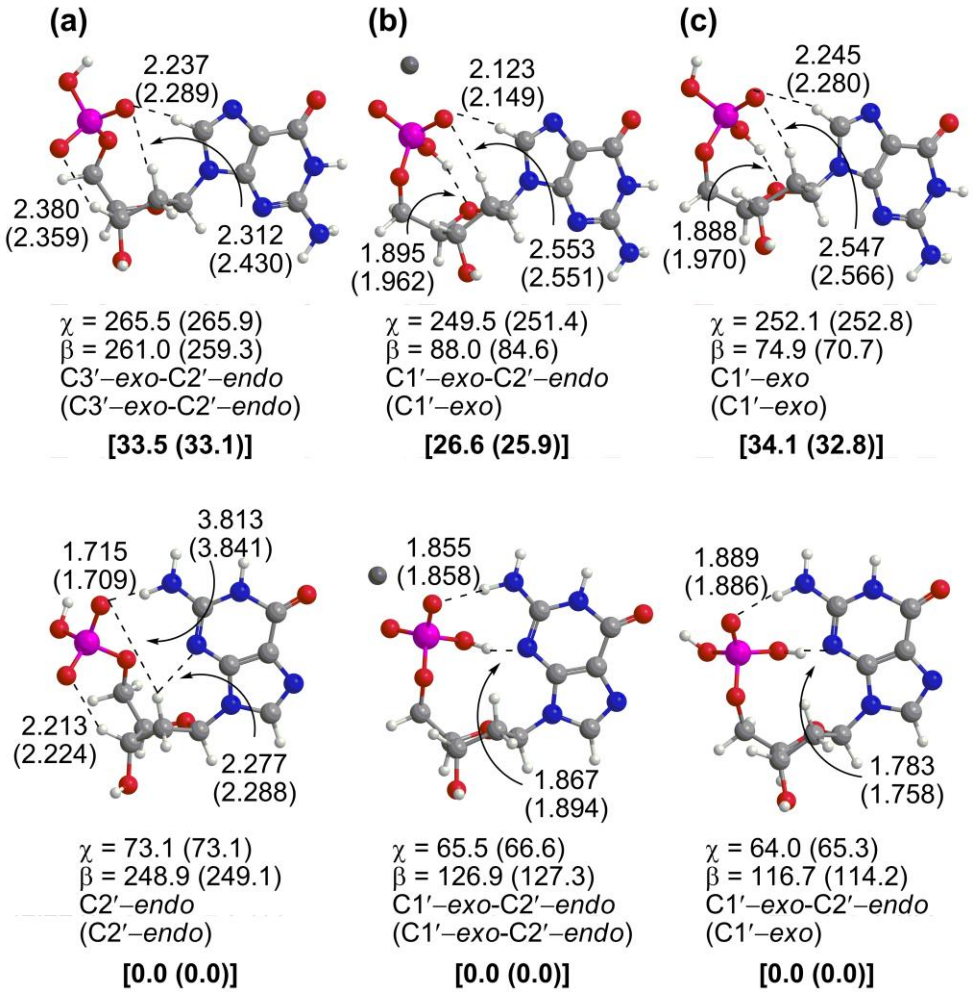


Figure 4.4The *anti* (top) and *syn* (bottom) gas-phase structures of natural dGMP described by the (a) anionic HPO_4^-, (b) counterion Na^+ HPO_4^-, and (c) neutral H_2PO_4 phosphate models. Select B3LYP/6-31G(d) hydrogen-bonds (Å), dihedral angles (χ and β, degrees), and sugar puckering are provided, where values for B3LYP/ 6-31+G(d,p) structures are provided in parentheses. B3LYP/6-311+G(2df,p) relative energies (kJ mol^{-1}) are in bold square brackets.

sugar intramolecular interactions distort the B-DNA (C2′–*endo*) sugar puckering, where the anionic model exhibits a C3′–*exo*-C2′–*endo* twist, and the neutral and counterion models adopt a C1′–*exo* pucker.

For all three *syn* dGMP models, the N2 amino group forms a hydrogen bond with an unprotonated phosphate oxygen (Figure 4.4, bottom), which is strongest (shortest) in the

anionic model followed by the counterion model, as expected based on the relative charge on the phosphate in each model. This contact is expected in B-DNA, and has been proposed to cause the *syn* conformation to be more stabilized for guanine than the other natural bases.[5,19,27,28] The nucleoside model does not contain this native interaction, which emphasizes the importance of considering larger, more complete nucleotide models. As discussed for the *anti* conformation, the counterion and neutral models share similar χ values (64 – 66°), while the anionic model differs by 8 – 9°. Consequently, the neutral and counterion models form a similar non-native hydrogen bond between the protonated phosphate oxygen and N3 of guanine (Figure 4.4b, c bottom), which is significantly shorter in the neutral model due to a 10° larger β angle. In contrast to these models, β (249°) rotates the phosphate group away from the nucleobase in the anionic model, which results in a hydrogen bond between the unprotonated phosphate oxygen and C3′–H of the sugar that is not expected in DNA.[27]

As previously found in the literature[11-13,27,30] calculated relative energies for all nucleotide models predict a preferred *syn* conformation for dGMP (Figure 4.4). The *syn* preference is large, ranging from 26.6 – 34.1 kJ mol^{-1}. Although the calculated *syn* stability is similar for the neutral and anionic model, it is smaller for the counterion model. The same result is found with the larger (6-31+G(d,p)) basis set, where the difference in the *syn* preference between basis sets is less than 2 kJ mol^{-1}. Furthermore, for both basis sets, the *syn* geometry is distorted from the biologically-relevant structure by the formation of non-native intramolecular hydrogen bonds. Therefore, these results demonstrate that gas-phase optimizations with these (double-zeta) basis sets do not yield an *anti*/*syn* energy trend or structures relevant to B-DNA.

Other computational studies have reported that including the effects of the (implicit) water environment as an energy correction decreases the relative energy of the

anti conformations of the natural nucleotides (but *syn* is still the preferred conformation).[13] Although inclusion of an energy correction alone cannot fix the structural problems reported here, these previous results suggest that including the effects of solvation directly in the optimization routine may yield different minimum energy structures. Therefore, the effects of solvation on the structure of the nucleotide are considered in the next section by performing optimizations in a (bulk) water environment.

4.3.2.2 Effects of the Water Environment on the Structure of Natural dGMP

The *anti* and *syn* conformers described by the three phosphate models were optimized in (implicit) water without imposing constraints on the backbone using the PCM continuum solvation method with the 6-31G(d) and 6-31+G(d,p) basis sets (Figure 4.5). Water was chosen as the solvent since the gas-phase structures cannot correctly model the phosphate orientations in DNA helices and the phosphate backbone is highly solvated in DNA. The resulting *anti* structures for the three models (Figure 4.5, top) differ significantly from the corresponding gas-phase conformations (Figure 4.4, top). While the χ values deviate by 40° in the anionic model and 5° in the neutral and counterion models, the β angles are shifted (by ~90°) to the range expected for B-DNA (~185°).[16] Importantly, no interactions between the non-native terminal 5′-phosphate hydrogen and the base or sugar moiety are observed. Indeed, overall there are fewer and weaker intramolecular hydrogen bonds in the *anti* structures optimized in water compared to the gas-phase geometries. These results indicate that, in a polar environment, intramolecular interactions between the phosphate and sugar or base are not a significant source of stabilization. Additionally, unlike the gas phase, the *anti* conformation optimized in water has B-DNA sugar puckering (C2′–*endo*).

Similar to the gas phase, a distorted *syn* conformation was optimized in water for all models (Figure 4.5, middle) with the phosphate group rotated toward the nucleobase (β =

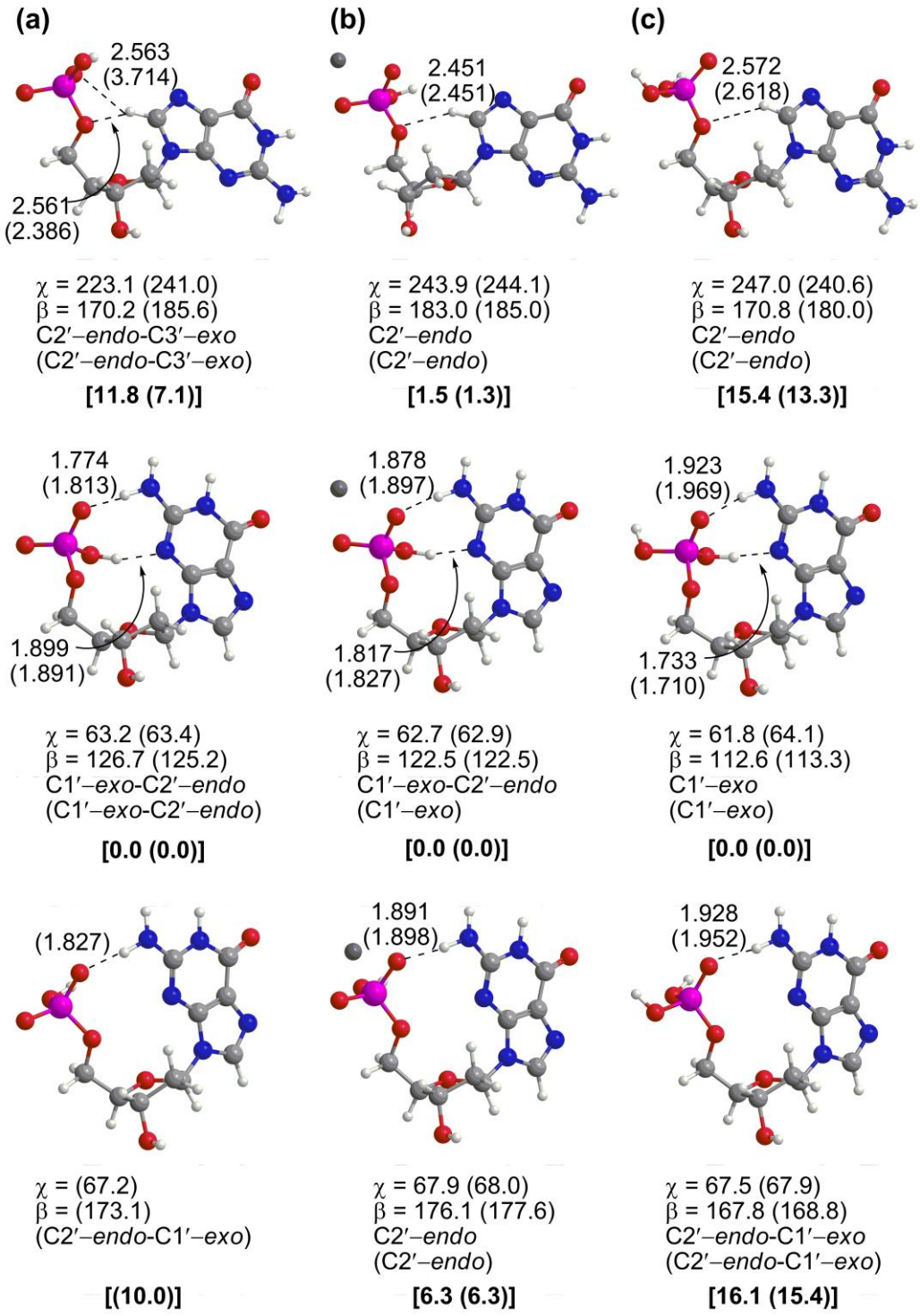


Figure 4.5 The *anti* (top), distorted *syn* (middle), and biologically-relevant *syn* (bottom) structures optimized in water for natural dGMP described by the (a) anionic HPO_4^-, (b) counterion Na^+ HPO_4^-, and (c) neutral H_2PO_4 phosphate models. Select B3LYP/6-31G(d) hydrogen-bonds (Å), dihedral angles (χ and β, degrees), and sugar

puckering are provided, where values for B3LYP/6-31+G(d,p) structures are provided in parentheses. B3LYP/6-311+G(2df,p) relative energies (kJ mol^{-1}) are in bold square brackets.

113 – 126°), and χ equal to 62 – 63°. Two hydrogen-bonding contacts are present in all three structures. First, a hydrogen bond between the N2 amino group of guanine and the phosphate oxygen was found. Although this interaction is also present in the gas phase, all three models have weaker (longer) bonds in water than in the gas phase. Second, a hydrogen-bond contact between the protonated phosphate oxygen and N3 of guanine is observed. Although this interaction was previously found for the counterion and neutral models (not the anionic model) in the gas phase, this interaction is stronger in water. This non-native hydrogen-bonding interaction renders these structures irrelevant to DNA.

In addition to the distorted *syn* structures that are similar to those optimized in the gas phase, the *syn* structures of biological interest were characterized in PCM water, where the phosphate backbone orientation resembles that typically found in B-DNA[16] (Figure 4.5, bottom). Specifically, the β values range from 168 to 176° and χ equals 67 – 68°. These geometrical parameters are also in good agreement with the previous large basis set gas-phase study.[27] However, a modified anionic model was implemented in this previous study that avoids intramolecular interactions by protonating a phosphate oxygen not involved in the phosphodiester linkage to the neighboring nucleotide in B-DNA. Similar to the *anti* (water) conformations, no unnatural hydrogen bonds between the protonated phosphate oxygen and the nucleobase are found in the *syn* (water) structures. Thus, for the first time, a biologically-relevant *syn* structure was characterized in the present work in the absence of constraints using a reasonably-sized basis set suitable for studying bulky DNA adducts. The interaction between the amino group and the phosphate oxygen that is expected in DNA[5,19,27,28] is still present in the models. Since this interaction is responsible for stabilizing

the *syn* conformation of the nucleotide, this again highlights the importance of considering nucleotide models when attempting to understand the conformational preference in DNA. It should also be noted that the N2–H···O–P hydrogen bond in the present study is weaker (1.8 – 1.9 Å) compared to the similar *syn* structure characterized in a previous gas-phase study (1.618 Å) using a modified anionic model.[27]

As mentioned above, the structures of the natural nucleotides were studied with two basis sets (6-31G(d) and 6-31+G(d,p)). The geometries obtained with both basis sets generally deviate by up to 10° with the exception of two significant differences, which both occur for the anionic water model. First, the *anti* anionic structures have different intramolecular hydrogen bonds since χ deviates by 18° and β by 15°. For the smaller basis set, this leads to a bifurcated hydrogen bond between C8–H and a phosphate oxygen and O5′ (Figure 4.5a, top), while the larger basis set results in a more direct hydrogen bond between C8–H and O5′. The second major difference is that no biologically-relevant *syn* structure is obtained for the anionic model (in water) with the smaller (6-31G(d)) basis set. This stresses the importance of including diffuse functions in the basis set in order to accurately describe charged models. Furthermore, this discrepancy suggests that the anionic model cannot be reliably used in instances where the size of the nucleobase prohibits the use of a larger basis set. Thus, the anionic model may not be suitable for studying bulky DNA adducts.

In addition to characterizing biologically-relevant *syn* conformations, consideration of solvated structures changes important conclusions regarding the *anti/syn* energy trend. Specifically, for all three models, the non-biologically relevant *syn* conformation that is similar to the gas-phase structure remains the global minimum by 1 – 15 kJ mol^{-1} (Figure 4.5). However, when the nucleotide conformation in DNA is of interest, this structure must be discarded and the most fair *anti*/*syn* energy difference is obtained by considering the

biologically-relevant *syn* minimum, which is only characterized in solvent-phase (water) optimizations.

At the B3LYP/6-31+G(d,p) level of theory, the biologically-relevant *syn* structure is higher in energy than the *anti* conformation by 2 – 5 kJ mol^{-1} for all three models (Figure 4.5, parentheses). Therefore, all three models correctly predict that *anti* is the preferred conformation about the glycosidic bond. At the B3LYP/6-31G(d) level of theory, the neutral and counterion models also predict the *anti* conformation to be lower in energy than the *syn* conformation by 0.7 and 4.8 kJ mol^{-1}, respectively. This is the first time the correct *anti*/*syn* energy trend has been computationally reproduced using an unconstrained nucleotide model for natural dGMP. It should be noted that the exaggerated N2–H···O–P interaction partially explains why previous work predicted the incorrect *anti*/*syn* energy trend,[27] where this error could have been caused by the basis set and/or the charged (anionic) model previously implemented. Although the structure in the present study is similar to the previous gas-phase structures,[27,30] small differences in the geometric parameters (i.e., hydrogen-bond lengths) are clearly important for determining the correct *anti*/*syn* preference. Therefore, the solvated structures provide a more accurate representation of dGMP than has been previously reported in the literature. The results stress the importance of including the effects of the (bulk) environment directly in the optimization routine, and that this approach has the same effect as adding geometric constraints, while still permitting a free optimization. This new approach will be particularly useful when studying damaged or modified bases with unknown geometries.

Although the larger (smaller) basis set correctly predicts the *anti* conformer to be lower in energy than *syn* for all three (two) of the phosphate models considered in the present work, there are important discrepancies in the *anti/syn* energetic differences that differentiate the models. First, the *anti* preference predicted by both basis sets is largest for

the counterion model. Second, the preference for the hydrogen-bonded *syn* structure calculated by both basis sets is smallest for the counterion model. Also, the biologically-relevant *syn* conformation is most readily optimized (i.e., has fewer convergence issues) when the counterion model is implemented with either basis set. The fact that the counterion model more reliably identifies a biologically-relevant *syn* structure for the natural nucleotide with both basis sets suggests that this model should be implemented when modeling DNA nucleotides. This conclusion is especially true when smaller basis sets are implemented for computational efficiency. Therefore, due to the size of bulky DNA adducts, future studies of these nucleotides in geometries relevant to B-DNA should be conducted using the counterion model and incorporating bulk solvent effects in the optimization routine.

4.3.2.3 Structure of the *o*-PhOH-dG Nucleotide Adduct

Since it was shown that the counterion model more reliably reproduces biologically-relevant structures and the *anti/syn* energetic preference for natural dGMP, this model will be employed to study the C-linked PhOH-dG DNA adducts. However, since the majority of work in the literature uses neutral and anionic nucleotide models,[4,5,9-14,18,19,27,28,30,31] these models will also be applied to the C-linked PhOH-dG adducts to further illustrate why they should not be used to study damaged nucleotide conformations. The *o*-PhOH-dGMP adduct was also optimized with both basis sets (6-31G(d) and 6-31+G(d,p)) to verify that the basis set effects are also small for damaged nucleotides, where interactions between the bulky group and the phosphate could occur. Indeed, the differences between the two basis sets are negligible (Figure 4.6), and therefore only the results for the 6-31G(d) basis set will be discussed below unless otherwise indicated. Due to the clear importance of directly including (bulk) environmental effects in the optimization routine, all structures of the DNA adducts were optimized in PCM water.

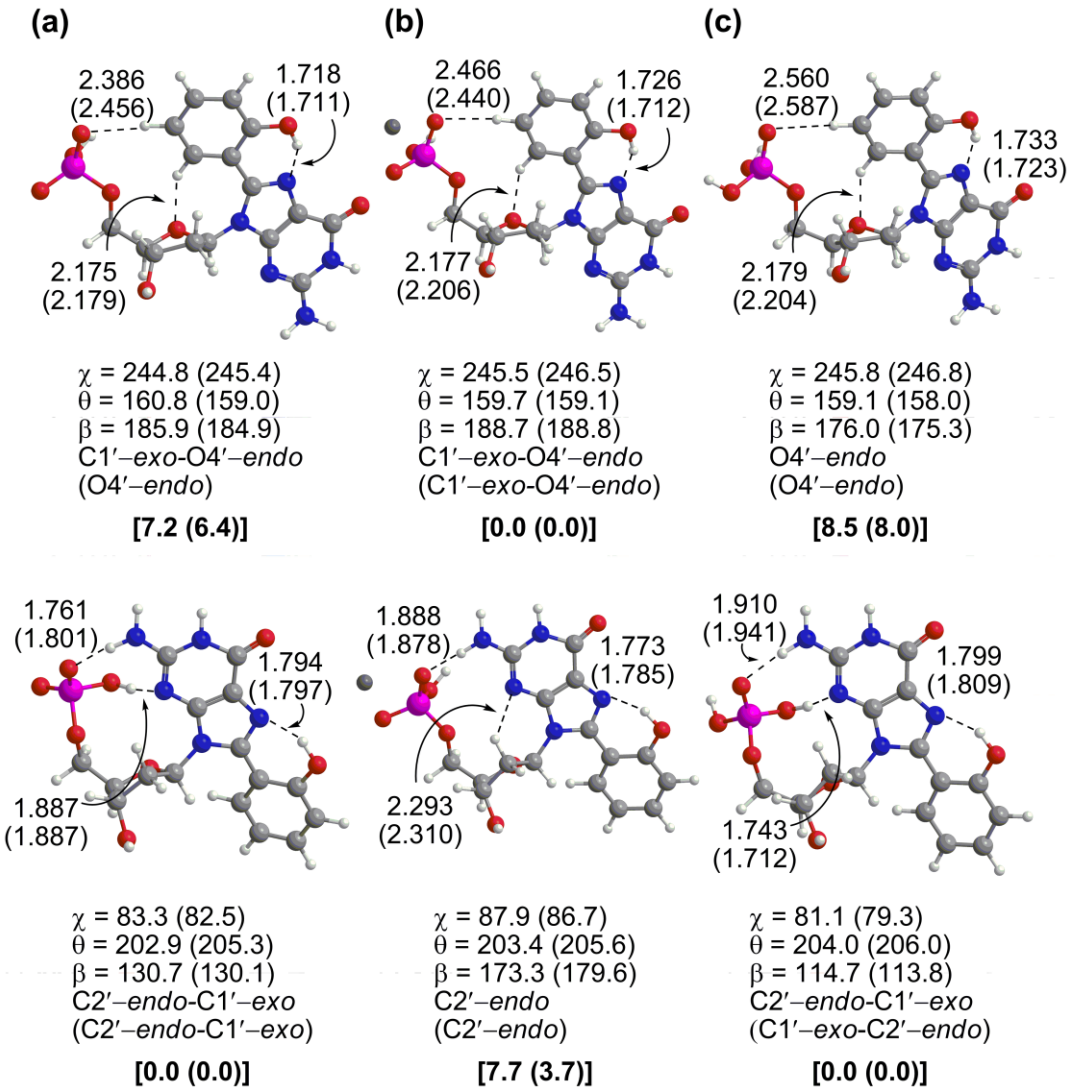


Figure 4.6 The *anti* (top) and *syn* (bottom) conformations of *o*-PhOH-dGMP described by the (a) anionic HPO_4^-, (b) counterion Na^+ HPO_4^-, and (c) neutral H_2PO_4 phosphate models. Select B3LYP/6-31G(d) hydrogen bonds (Å), dihedral angles (χ and β, degrees), and sugar puckering are provided, where values for B3LYP/6-31+G(d,p) structures are provided in parentheses. B3LYP/6-311+G(2df,p) relative energies (kJ mol^{-1}) are in bold square brackets.

The optimized *anti* structures of *o*-PhOH-dGMP with all models are very similar (Figure 4.6, top). The χ dihedral angle (~245°) is close to the calculated values for the natural nucleotide (223 – 247°). Similarly, the β-dihedral angle varies only slightly across models (176 – 186°), and is shifted by only 5° in the neutral and counterion model and 15°

in the anionic model compared with the natural nucleotide. Interestingly, the bulky group forms a weak C14–H···O hydrogen bond with the phosphate group in the *anti* conformation, which is similar to the C8–H···O interaction discussed in the *anti* natural nucleotide (Figure 4.6, top). A second weak C15–H···O hydrogen bond also forms with the sugar. This leads to the only significant difference in the *anti* structures, which is a change in the sugar puckering to become more O4′–*endo* upon damage. This interaction with the phosphate group represents a significant difference between the nucleoside and nucleotide models, which re-emphasizes the importance of considering larger models to understand DNA structure where phosphate interactions may become important in the helical environment.

In the *syn* structures of *o*-PhOH-dGMP (Figure 4.6, bottom), the sugar puckering (C2′–*endo* or C2′–*endo*-C1′–*exo*) remains approximately the same as for the natural nucleotide. Each model also exhibits a hydrogen bond between the N2 guanine amino group and the phosphate oxygen, which is much stronger than the C–H···O interactions observed in the *anti* conformation. Nevertheless, there are some major differences between the models. Specifically, although χ increases by 20° and β increases by 2 – 4° relative to the natural *syn* nucleotide for the neutral and anionic models, these models only characterize the non-native, hydrogen-bonded *syn* conformation discussed for the natural nucleotide in both the gas-phase and water. Indeed, the biologically-relevant *syn* structure was not optimized with the anionic and neutral models using either basis set. However, the counterion model readily characterizes the biologically-relevant *syn* structure with both basis sets, which involves deviations in β and χ by 3° and 20°, respectively, compared to the corresponding natural conformation. Furthermore, the hydrogen-bonded *syn* structure was not found for the counterion model. Thus, comparison of the structures optimized with all three models reinforces the previous conclusion that the counterion model is the best choice for modeling natural and damaged B-DNA nucleotides. In addition, the agreement

between both basis sets confirms that *p*-PhOH-dGMP can be studied at only the B3LYP/ 6-31G(d) level of theory.

Although the neutral and anionic models predict the non-native hydrogen-bonded *syn* conformation to be energetically preferred as discussed for the natural nucleotide, the calculated energy difference between the *anti* and non-native *syn* conformations is smaller for *o*-PhOH-dG (Figure 4.6). More importantly, the biologically-relevant *syn* structure isolated using the counterion model is less stable than the *anti* conformation as discussed for the natural nucleotide. Interestingly, the energetic preference for the *anti* conformer is larger for *o*-PhOH-dG (7.7 kJ mol^{-1}) than the natural nucleotide (4.8 kJ mol^{-1}). This suggests that *o*-PhOH-dG may adopt the *anti* conformation in B-DNA, and therefore *o*-PhOH-dG should behave like natural dG in terms of the base-pairing preference in DNA helices,[41] as discussed in Chapter 3. However, the predicted preference for *anti* is still small, and other environmental factors in the helix (such as stacking interactions) may influence the preferred conformation.

4.3.2.4 Structure of the *p*-PhOH-dG Nucleotide Adduct

As discussed above, the structural differences for any given conformation of *o*-PhOH-dGMP obtained with the two basis sets are less than 6° (Figure 4.6) and the same conclusions are obtained regardless of the basis set implemented. Consequently, the *p*-PhOH-dG nucleotide adduct was only studied using the smaller (6-31G(d)) basis set. The *anti* structures of *p*-PhOH-dGMP optimized with all models are very similar (Figure 4.7, top). The χ dihedral angle (~250°) is shifted by only 5° compared with *o*-PhOH-dGMP, while the β-dihedral angles (175 – 185°) and sugar puckering (O4′–*endo*-C1′–*exo*) are identical for both adducts. A C14–H···O interaction between the bulky phenoxyl group and the phosphate is present, which again stresses the importance of considering nucleotide models. As discussed for *o*-PhOH-dGMP, only the non-native hydrogen-bonded *syn* structure

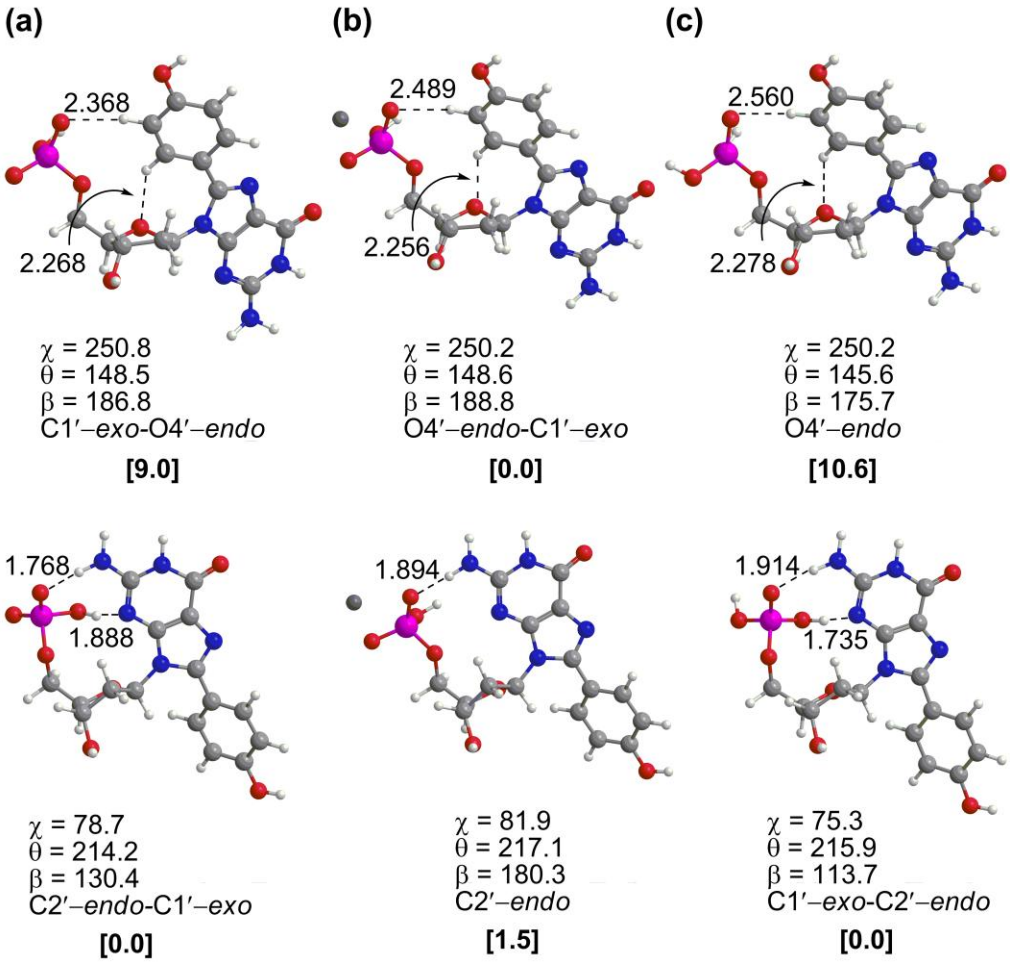


Figure 4.7 The *anti* (top) and *syn* (bottom) conformations of *p*-PhOH-dGMP described by the (a) anionic HPO_4^-, (b) counterion Na^+ HPO_4^-, and (c) neutral H_2PO_4 phosphate models. Select B3LYP/6-31G(d) hydrogen-bonds (Å), dihedral angles (χ and β, deg.), and sugar puckering are provided. B3LYP/6-311+G(2df,p) relative energies (kJ mol^{-1}) are in bold square brackets.

was characterized using the neutral and anionic models, while the biologically-relevant *syn* conformation is only obtained for the counterion model. This finding once again recommends using this model in future studies. The *p*-PhOH-dGMP structures predicted with the counterion model differ from the *o*-PhOH-dGMP structures by only 7° in the β angle and 5° in χ, but *p*-PhOH-dGMP lacks the interaction between N3 and C2′–H noted for *o*-PhOH-dGMP.

As discussed for the *o*-PhOH-dG nucleotide adduct, the biologically-relevant *syn* structure isolated using the counterion model leads to a preference for the *anti* conformation for *p*-PhOH-dGMP (Figure 4.7). However, while the energetic preference for *anti* is greater for *o*-PhOH-dGMP (7.7 kJ mol^{-1}) than the natural nucleotide (4.8 kJ mol^{-1}), the preference for *anti* is actually less for *p*-PhOH-dGMP (1.5 kJ mol^{-1}). Therefore, the *anti* conformation is destabilized for *p*-PhOH-dGMP relative to the natural base, and the very small energy difference suggests that the *syn* conformation could be adopted in B-DNA helices. Combined with the differences in base-pair stabilities for the *o*-PhOH-dG and *p*- PhOH-dG adducts reported in Chapter 3, this implies that the preferred nucleobase conformation in C8-phenoxyl nucleotide adducts in B-DNA duplexes may depend on the relative location of the hydroxyl substituent.

4.4 Conclusions

The *anti/syn* conformational preference of the *o*- and *p*-PhOH-dG nucleotide adducts was investigated in this Chapter. Similar to previous studies of natural dG, fully-optimized bulky adduct nucleoside models predict the *syn* conformation to be favored due to the presence of a strong O5′–H···N3 hydrogen bond, which cannot occur in DNA. When an additional geometric constraint is imposed on the C5'-hydroxyl group that prevents this interaction, the *syn* conformation becomes less important than initially predicted. Furthermore, when this constraint is released in subsequent optimizations, the resulting *anti* structures are nearly thermoneutral with the original *syn* global minimum. Therefore, although the nucleoside model is popular in the literature and can provide some useful information regarding the structure of damaged bases, this model is insufficient to determine the *anti*/*syn* conformational preference of modified nucleobases in DNA.

To better predict the structure of the damaged bases in DNA helices, a nucleotide model that includes the 5′-monophosphate group was studied. However, since previous

work could not predict the correct *anti*/*syn* preference for natural dGMP with a computational methodology suitable for bulky adducts, dGMP was first considered using both gas-phase and solvent-phase optimizations and three computational models for the phosphate group (anionic HPO_4^-, neutral H_2PO_4, or (neutral) counterion Na^+ HPO_4^-). Structures obtained using these models significantly differ from the results obtained with the nucleoside models, which emphasizes the importance of using nucleotide models for applications to DNA. The results confirm that biologically-relevant *syn* conformations, which are void of non-native hydrogen-bonding interactions, are characterized only when the effects of the (bulk) environment are taken into account during the optimization routine. Most importantly, this *syn* conformation is predicted to be less stable than the *anti* conformation, which agrees with the preference in B-DNA helices. Since the preference for the *anti* conformation is greatest when the counterion model is implemented (regardless of the basis set employed), the counterion model is concluded to be the most reliable for studying DNA nucleotides. Although the calculated energy difference is small (2 – 5 kJ mol^{-1}), this was the first time the correct conformational trend was calculated for natural dGMP without the use of artificial geometric constraints.

The methodology used for the natural nucleotide was subsequently applied to the damaged nucleotides to determine whether the *o*- and *p*-PhOH-dGMP adducts share the natural preference for *anti* or whether they adopt a *syn* conformation, which would suggest a greater potential for mutagenicity. Although both types of damage show a weak preference for the *anti* conformation, the *o*-PhOH-dGMP adduct has a stronger preference for the *anti* conformation than the natural nucleotide, while the *p*-PhOH-dGMP adduct has the smallest *anti*/*syn* energy difference. This suggests that the preferred structure of bulky nucleotide adducts in DNA may depend on small structural differences (i.e., the location of the hydroxyl substituent on the phenyl group). This hypothesis is supported by the

differences in base-pair stabilities for the adducts discussed in Chapter 3, where *p*-PhOH-dG formed more stable base pairs in the *syn* conformation than *o*-PhOH-dG. However, other sources of stabilization (such as intrastrand interactions or steric clashes) may play an important role in determining the adduct structure in damaged DNA helices, where a complex conformational heterogeneity may exist depending on the sequence context of the adduct. Therefore, the following chapter considers intrastrand interactions with flanking bases using a deoxydinucleoside monophosphate model.

4.5 References

(1) Hocquet, A.; Leulliot, N.; Ghomi, M. *J. Phys. Chem. B* **2000**, *104*, 4560-4568.

(2) Foloppe, N.; MacKerell, A. D. *J. Phys. Chem. B* **1999**, *103*, 10955-10964.

(3) Shishkin, O. V.; Pelmenschikov, A.; Hovorun, D. M.; Leszczynski, J. *J. Mol. Struct.* **2000**, *526*, 329-341.

(4) Foloppe, N.; Nilsson, L.; MacKerell, A. D. *Biopolymers* **2001**, *61*, 61-76.

(5) Foloppe, N.; Hartmann, B.; Nilsson, L.; MacKerell, A. D. *Biophys. J.* **2002**, *82*, 1554-1569.

(6) Schlitt, K. M.; Sun, K. W. M.; Paugh, R. J.; Millen, A. L.; Navarro-Whyte, L.; Wetmore, S. D.; Manderville, R. A. *J. Org. Chem.* **2009**, *74*, 5793-5802.

(7) McLaughlin, C. K.; Lantero, D. R.; Manderville, R. A. *J. Phys. Chem. A* **2006**, *110*, 6224-6230.

(8) Millen, A. L.; McLaughlin, C. K.; Sun, K. M.; Manderville, R. A.; Wetmore, S. D. *J. Phys. Chem. A* **2008**, *112*, 3742-3753.

(9) Kosenkov, D.; Kholod, Y. A.; Gorb, L.; Shishkin, O. V.; Kuramshina, G. M.; Dovbeshko, G. I.; Leszczynski, J. *J. Phys. Chem. A* **2009**, *113*, 9386-9395.

(10) Shishkin, O. V.; Gorb, L.; Zhikol, O. A.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2004**, *21*, 537-553.

(11) Shishkin, O. V.; Gorb, L.; Zhikol, O. A.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2004**, *22*, 227-243.

(12) Gorb, L.; Shishkin, O.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2005**, *22*, 441-454.

(13) Palamarchuk, G. V.; Shishkin, O. V.; Gorb, L.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2009**, *26*, 653-661.

(14) , A.; Cheatham III, T. E.; , J. i. *J. Phys. Chem. B* **2008**, *112*, 8188-8197.

(15) Kosenkov, D.; Gorb, L.; Shishkin, O. V.; Sponer, J.; Leszczynski, J. *J. Phys. Chem. B* **2008**, *112*, 150-157.

(16) The β angle in B-DNA is typically 168 ± 25°. (See Reference 4.)

(17) Sarma, R. H.; Lee, C. H.; Evans, F. E.; Yathindr.N; Sundaral.M *J. Am. Chem. Soc.* **1974**, *96*, 7337-7348.

(18) MacKerell, A. D. *J. Phys. Chem. B* **2009**, *113*, 3235-3244.

(19) Yathindra, N.; Sundaralingam, M. *Biopolymers* **1973**, *12*, 297-314.

(20) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502-16513.

(21) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03, Revisions C.02 and D.01; Gaussian, Inc.: Wallingford, CT, 2004.

(22) Although four *anti* structures can be identified on the potential energy surface for the *para* adduct (Figure 4.2, right), only three unique structures were characterized following full optimizations. Specifically, the *anti* structure corresponding to $\chi \sim 200°$ and $\theta \sim 40°$ optimizes to $\chi \sim 250°$ and $\theta \sim 330°$.

(23) The *para* structure corresponding to $\chi \sim 60°$ and $\theta \sim 140°$ and the *ortho* structures corresponding to $\chi \sim 80°$, $\theta \sim 60°$ and $\chi \sim 60°$, $\theta \sim 310°$ optimized to a C1′-*exo*-C2′-*endo* twist sugar puckering.

(24) Three *anti* conformations significantly deviate from the natural b angle such that the hydroxyl group is above O4′ of the sugar (*para* with $\chi \sim 220°$ and $\theta \sim 210°$, and *ortho* with $\chi \sim 190°$ and $\theta \sim 70°$ or $\chi \sim 250°$ and $\theta \sim 300°$). Since structures where the hydroxyl group rotated toward the nucleobase remain higher in energy than structures relevant to DNA, it is not expected these would be lower in energy if the hydroxyl group had remained in the natural orientation.

(25) Chen, X.; Zhan, C. G. *J. Phys. Chem. B* **2008**, *112*, 16851-16859.

(26) Poltev, V. I.; Anisimov, V. M.; Danilov, V. I.; Deriabina, A.; Gonzalez, E.; Jurkiewiez, A.; Les, A.; Polteva, N. *J. Biomol. Struct. Dyn.* **2008**, *25*, 563-571.

(27) Shishkin, O. V.; Palamarchuk, G. V.; Gorb, L.; Leszczynski, J. *J. Phys. Chem. B* **2006**, *110*, 4413-4422.

(28) Yathindra, N.; Sundaralingam, M. *Biopolymers* **1973**, *12*, 2075-2082.

(29) Leulliot, N.; Ghomi, M.; Scalmani, G.; Berthier, G. *J. Phys. Chem. A* **1999**, *103*, 8716-8724.

(30) Rubio, M.; Roca-Sanjuan, D.; Serrano-Andres, L.; Merchan, M. *J. Phys. Chem. B* **2009**, *113*, 2451-2457.

(31) Zakjevskii, V. V.; Dolgounitcheva, O.; Zakrzewski, V. G.; Ortiz, J. V. *Int. J. Quantum Chem.* **2007**, *107*, 2266-2273.

(32) Wiechelman, K.; Taylor, E. R. *J. Biomol. Struct. Dyn.* **1998**, *15*, 1181-1194.

(33) Foloppe, N.; Nilsson, L. *J. Phys. Chem. B* **2005**, *109*, 9119-9131.

(34) Foloppe, N.; MacKerell, A. D. *Biophys. J.* **1999**, *76*, 3206-3218.

(35) Yurenko, Y. P.; Zhurakivsky, R. O.; Ghomi, M.; Samijlenko, S. P.; Hovorun, D. M. *J. Phys. Chem. B* **2007**, *111*, 6263-6271.

(36) For recent studies that used nucleoside models see, for example, (a) Chen, Z. Q.; Zhang, C. H.; Xue, Y. *J. Phys. Chem. B* **2009**, *113*, 10409-10420. (b) Sun, L. X.; Bu, Y. X. *Theochem-J. Mol. Struct.* **2009**, *909*, 25-32. (c) Yang, Z. Z.; Qi, S. F.; Zhao, D. X.; Gong, L. D. *J. Phys. Chem. B* **2009**, *113*, 254-259. (d) Tehrani, Z. A.; Fattahi, A.; Pourjavadi, A. *Carbohydr. Res.* **2009**, *344*, 771-778. (e) Barbe, S.; Le Bret, M. *J. Comput. Chem.* **2008**, *29*, 1353-1363. (f) O'Daniel, P. I.; Jefferson, M.; Wiest, O.; Seley-Radtke, K. L. *J. Biomol. Struct. Dyn.* **2008**, *26*, 283-292. (g) Yurenko, Y. P.; Zhurakivsky, R. O.; Ghomi, M.; Samifflenko, S. P.; Hovorun, D. M. *J. Phys. Chem. B* **2008**, *112*, 1240-1250. (h) Rios-Font, R.; Rodriguez-Santiago, L.; Bertran, J.; Sodupe, M. *J. Phys. Chem. B* **2007**, *111*, 6071-6077.

(37) Gelbin, A.; Schneider, B.; Clowney, L.; Hsieh, S. H.; Olson, W. K.; Berman, H. M. *J. Am. Chem. Soc.* **1996**, *118*, 519-529.

(38) Altona, C.; Sundaral.M *J. Am. Chem. Soc.* **1973**, *95*, 2333-2344.

(39) Chachaty, C.; Perly, B.; Forchioni, A.; Langlet, G. *Biopolymers* **1980**, *19*, 1211-1239.

(40) Vokacova, Z.; Budesinsky, M.; Rosenberg, I.; Schneider, B.; Sponer, J.; Sychrovsky, V. *J. Phys. Chem. B* **2009**, *113*, 1182-1191.

(41) The relative energies obtained with B3LYP/6-31+G(d,p) geometries for natural dG and the *ortho* adduct are only 1 kJ mol^{-1} different.

5 Chapter 5: Conformational Flexibility of C-linked PhOH-dG Adducts Using a Deoxydinucleoside Monophosphate Model[a]

5.1 Introduction

The previous chapter reveals that the *anti* and *syn* conformations of the dG nucleoside adducts are nearly thermoneutral, while nucleoside 5′-monophosphate adducts have a weak preference for the *anti* conformation. However, the calculated *anti/syn* energy difference is small regardless of model used, which suggests that both conformations may exist within DNA helices. Since other factors within duplexes may be able to preferentially stabilize one conformation, increasing the model size may further clarify the conformational preferences of the adducts. For example, the identity of the flanking bases has influenced the preferred conformation of other bulky C8 adducts.[1-5] Therefore, the current chapter expands the computational models of *o*-PhOH-dG and *p*-PhOH-dG by considering damaged deoxydinucleoside monophosphates (Figure 5.1) with an adduct in either the 5' or 3' position and each of the four natural nucleobases (A, C, G, T) in the corresponding flanking position.

Since similar models have been successfully used to investigate the sequence dependence of the conformations of natural B-DNA,[6-8] a deoxydinucleoside monophosphate model will afford information about structures relevant to the DNA environment. Indeed, the effects of sequence context on the preferred *anti*/*syn* conformation of other bulky C8 adducts (AF-dG and AAF-dG) were similarly studied using deoxydinucleoside monophosphate models[9,10] prior to investigating the adducts in larger sequences.[11-15] The aforementioned studies on natural strands and those containing other bulky adducts

---

[a] Reprinted in part with permission from J. Phys. Chem. B, submitted for publication. Millen, A. L.; Kamenz, B. L.; Leavens, F. M. V.; Manderville, R. A.; Wetmore, S. D. Unpublished work copyright 2011 American Chemical Society. Manuscript ID: jp-2011-057332. B.L.K. performed some calculations on the *para* dinucleoside monophosphate systems and F.M.V.L. performed some calculations on the *ortho* dinucleoside monophosphate systems.

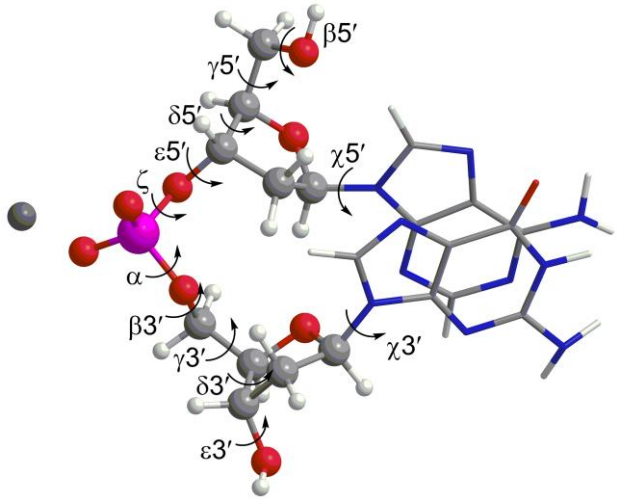


Figure 5.1 A representative deoxydinucleoside monophosphate with a sodium counterion phosphate model considered in the present study, and definition of important backbone dihedral angles (β5′ = ∠(HO5′ C5′ C4′), γ = ∠(O5′ C5′ C4′ C3′), δ = ∠(C5′ C4′ C3′ O3′), ε5′ = ∠(C4′ C3′ O3′ P), ζ = ∠(C3′ O3′ PO5′), α = ∠(O3′ PO5′ C5′), β3′ = ∠(PO5′ C5′ C4′) and ε3′ = ∠(C4′ C3′ O3′ H)). The (χ5′ and χ3′) glycosidic bond dihedral angles are equal to ∠(O4′ C1′ N9C4) for purine and ∠(O4′ C1′ N1C2) for pyrimidine-containing nucleotides.

suggest that this work will also permit the close examination of sequence effects on the conformational preference of the phenoxyl adducts, which are currently unknown. Indeed, experimental studies[16] discussed in Chapter 2 predicted that discrete interactions with solvent can disrupt the intramolecular O–H···N7 hydrogen bond and thereby increase the twist. This suggests that the surrounding environment may play an important role in determining the orientation of the bulky group, which in turn can influence the preferred base pairing as discussed in Chapter 3. Therefore, in addition to broadening the current knowledge of the conformational preference and sequence dependence of the adducts, the information gained from this study will be useful for identifying sequence contexts with interesting base–base interactions for study in large-scale molecular dynamics simulations of phenoxyl adducts incorporated into DNA helices.

5.2 Computational Details

Deoxydinucleoside monophosphate models were generated for base sequences with the natural, *o*- or *p*-PhOH-G nucleobase in the 5' or 3' position, and each of the natural bases in the corresponding flanking position (see, for example, Figure 5.1). Initially, the structure of the dApdG sequence was built from a central dinucleotide in a well-studied experimental DNA dodecamer crystal structure (PDB code: 436D)[17] and optimized according to the methodology outlined below. Subsequently, the nucleobase sequence was altered to generate all natural sequences examined in this chapter. Corresponding damaged sequences were generated by adding the phenoxyl substituent to G, and setting the sugar puckering and the θ (∠(N9C8C10C11)) dihedral angle to the values previously calculated using a dGMP model (Chapter 4). Subsequently, the natural or damaged guanine nucleobase was rotated about the glycosidic bond to generate the corresponding *syn* structures.

The computational literature studying DNA structure and reactions has used a variety of different phosphate models in attempts to describe the complicated environment of DNA helices.[6,18-29] Although anionic and neutral (protonated) phosphate models have been widely used, Chapter 4 shows that neither model accurately reproduces the *anti*/*syn* stability of the natural dGMP nucleotide. Therefore, the present study uses a counterion phosphate model where a sodium ion is equidistant between two oxygen atoms in the phosphate backbone. This model also predicts the correct *anti/syn* preference for dG in a nucleotide model (Chapter 4) and successfully reproduces the average values of B-DNA backbone dihedral angles in natural deoxydinucleoside monophosphates.[6] Additionally, this model likely provides a more accurate representation of the DNA environment, where counterions are common in solution.[6]

Each possible sequence of the deoxydinucleoside monophosphate with the sodium counterion phosphate model was optimized with M06-2X/6-31G(d,p) in water (ε = 78)

using the IEF-PCM formalism.[30] Although previous chapters used B3LYP to study the conformational properties of the corresponding nucleoside and nucleotide models, this functional incorrectly accounts for dispersion interactions.[31] Therefore, B3LYP is not a viable choice for studying deoxydinucleoside monophosphates, which include stacking interactions. In contrast, M06-2X was designed in part to yield more accurate noncovalent interactions that contain significant dispersion contributions, as well as reliable thermochemical data.[32] The effects of the (bulk) environment were taken into account during the optimization step since Chapter 4 concluded this approach is necessary to successfully optimize the biologically-relevant *syn* conformation of dGMP. Although a relatively small basis set is implemented in the present chapter due to the size of the models, the chosen functional, basis set and environment combination is further justified by the good agreement between the structure of natural deoxydinucleoside monophosphate models and B-DNA (see discussion below).

All geometric variables were freely optimized with the exception of the β5′ (∠(HOC5'C4'), Figure 5.1) dihedral angle, which was constrained to 180° to prevent interactions with the nucleobase that are non-native to DNA helices. Frequency calculations were also performed on all structures. Reported relative energies were obtained from M06-2X/6-311+G(2df,p) single-point calculations and include scaled (0.9580) zero-point vibrational energy corrections. [33] All calculations were performed with Gaussian 09.[34]

5.3 Results and Discussion

This section considers the effects of phenoxyl substituents on the structure, and particularly the *anti/syn* conformational preference, of deoxydinucleoside monophosphate sequences containing dG. All possible nucleobase sequences are considered with dG, *o*-PhOH-dG or *p*-PhOH-dG located at the 5′ or 3′ position and one of the four natural bases in the corresponding flanking position. The discussion of the optimized structures will

focus on the relative base–base orientation and the backbone conformation since bulky adducts have been shown to affect both of these geometric variables.[10,35,36] As done in previous work,[7,8] the relative base–base orientation will be assessed by measuring the angle (designated as φ) between the normal vector to the base planes (defined as the mean plane generated using endocyclic heavy atoms). φ combines the values of tilt and roll commonly used to describe relative nucleobase arrangements in DNA double helices,[37] where φ = 0 indicates a parallel stacked base–base orientation. The backbone conformation will be compared to natural DNA[38] by analysing important torsional angles along the sugar-phosphate backbone (Figure 5.1). Due to the large number of sequences considered and the number of interesting geometric features of DNA strands, select examples are used to illustrate the major findings. In the discussion below, the structure and *anti/syn* conformational preference for natural dG sequences will be initially considered, and subsequently the effects of DNA damage on these parameters will be examined for the *o*-PhOH-dG and *p*-PhOH-dG adducts.

5.3.1 Natural Deoxydinucleoside Monophosphates

Figure 5.2 displays representative structures of the natural deoxydinucleoside monophosphates. In all sequences, the nucleobases remain in an orientation relevant to natural DNA helices and do not exhibit significant distortion compared to crystal structure data.[38] More specifically, despite a small degree of tilt between the bases in some of the deoxydinucleoside monophosphates (Table 5.1), stacked base arrangements are obtained for all sequences with dG in the *anti* conformation regardless of whether located in the 5' or 3' position (see, for example, Figures 5.2 a and b). Indeed, as previously reported for a deoxydinucleoside monophosphate described using a counterion phosphate model,[7] φ is generally less than 12° except for sequences with a pyrimidine in the 5' position (φ ≈ 20 – 25°, Table 5.1). Although highly planar base–base conformations are observed when dG

adopts the *syn* conformation in the 5' position (1° < φ < 10°), a greater degree of tilt (φ = 16 – 27°) is found for *syn* dG in the 3′ position (Table 5.1). The tilted base–base arrangements observed in the latter structures maximize interactions between the guanine nucleobase and both the phosphate (N2–H···O) and sugar (C2'–H···π) groups (Figure 5.3a). Regardless, the relative base–base orientations reported here closely resemble those found in B-DNA helices, which exhibit variations in stacked base–base arrangements.[8]

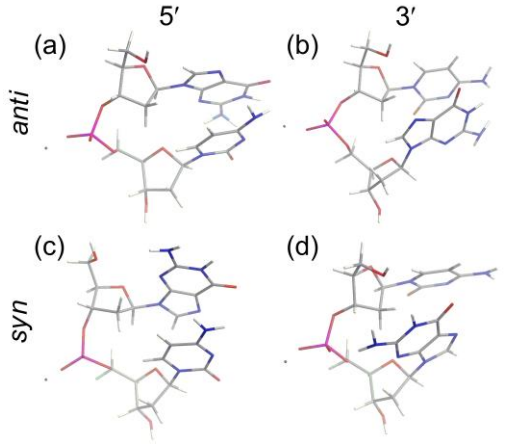


Figure 5.2 M06-2X/6-31G(d,p) structure of natural dGpdC (a, c) and dCpdG (b, d) deoxydinucleoside monophosphates with dG in the *anti* (a, b) or *syn* (c, d) conformation.

Since bulky DNA adducts may significantly affect the local structure of the DNA backbone,[36] it is particularly important to analyze the backbone orientation in the optimized natural sequences. This will ensure the model is accurate and permit reliable comparisons to the bulky adducts for which crystal structure data are scarce. Due to the number of configurations and backbone torsion angles considered (Figure 5.1), Table 5.2 displays the average values across all sequences according to the position (5′ or 3′) of *anti* or *syn* dG, while data for individual structures are provided in the Appendix (Tables A.1 – A.3, Appendix A). The *anti* χ values fall within the experimental averages obtained from

Table 5.1 Tilt angles defining the base–base orientation (φ) and relative arrangement of the PhOH and G Rings (θ) for deoxydinucleoside monophosphates containing natural dG, or a (*o*- or *p*-) PhOH-dG adduct[a, b]

Position & Conformer		Base	C		T		A		G	
			φ	θ	φ	θ	φ	θ	φ	θ
5′	*anti*	G	10.2	–	5.9	–	12.7	–	11.2	–
		para	15.9	144.6	15.5	150.3	2.5	148.4	22.1	144.7
		ortho	18.4	149.6	18.5	156.4	21.9	147.4	13.7	140.7
	syn	G	1.6	–	7.2	–	1.8	–	9.2	–
		para	7.6	217.7	10.2	215.9	4.1	218.1	14.2	223.3
		ortho	11.5	222.7	12.0	212.7	17.2	237.5	14.3	211.3
3′	*anti*	G	20.3	–	24.8	–	9.2	–	11.2	–
		para	21.0	146.4	23.1	139.9	18.6	127.4	28.3	142.6
		ortho	18.8	155.7	20.4	156.3	9.4	157.6	10.8	159.9
	syn	G	26.1	–	24.3	–	16.3	–	16.5	–
		para	27.3	219.8	24.0	212.2	25.5	207.4	22.6	212.6
		ortho	25.4	204.0	23.8	204.4	24.3	201.4	22.4	204.5

[a] M06-2X/6-31G(d,p) optimized geometries. φ is the degree of tilt between the base planes defined by the mean plane generated using endocyclic heavy atoms. θ (∠(N9C8C10C11) defines the degree of twist between the nucleobase and the bulky substituent (θ = 0 or 180° represents a planar structure). [b] For the *ortho* adduct nucleotide model, θ = 159.7° (*anti*) or 203.4° (*syn*). For the *para* adduct nucleotide model, θ = 148.6° (*anti*) or 217.1° (*syn*). (See Chapter 4.)

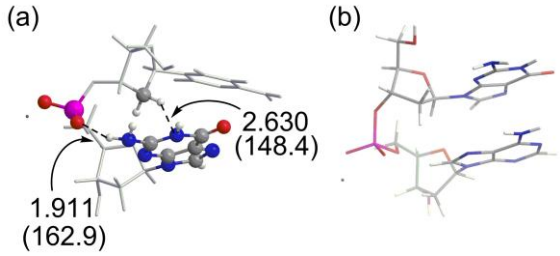


Figure 5.3 M06-2X/6-31G(d,p) structures [hydrogen-bond distances in Å (angles in degrees)] for *syn* dG in the (a) 3′ or (b) 5′ position.

crystal structures of B-DNA,[38] and are also very close to the values calculated using a nucleotide model (χ = 243.9°, Chapter 4). The remainder of the calculated dihedral angles are also in good agreement with those observed in B-DNA regardless of the location of the *anti* dG conformation (Table 5.2), and resemble previous results obtained using a similar computational approach.[6,7] The χ values of the *syn* conformation vary (by 4 – 17°) from

those reported for the nucleotide model (χ = 67.9°, Chapter 4). The remaining backbone torsion angles for deoxydinucleoside monophosphates containing *syn* dG at the 3' position deviate by less than 13° from the average crystal structure values for *anti* dG (Table 5.2), which suggests an *anti*/*syn* conversion has a small impact on the structure of the backbone. However, larger deviations from the average experimental backbone conformation occur when dG is in the 5′ position, where the variation is generally less than 10°, but can be up to 21 – 26° (see ε5′ and δ3′, Table 5.2). The larger deviations for strands with *syn* dG in the 5′ position suggest that the backbone may undergo distortion in order to accommodate the *syn* conformation while maintaining a planar base–base arrangement (Figure 5.3b).

Table 5.2 Average values of backbone torsional angles (degrees) in deoxydinucleoside monophosphates containing natural dG[a]

	Exp[b]	dG			
		anti		*syn*	
		5′	3′	5′	3′
χ5′	–[c]	248.1	239.1	84.9	231.5
χ3′	–[c]	248.0	274.8	254.6	72.2
γ5′	48 ± 11	48.6	50.2	48.4	48.3
δ5′	128 ± 13	147.6	145.2	140.9	140.7
ε5′	184 ± 11	190.0	194.6	158.0	190.6
ζ	265 ± 10	276.9	279.0	266.5	276.6
α	298 ± 15	294.8	295.7	301.7	300.9
β3′	176 ± 9	166.8	170.2	182.8	171.0
γ3′	48 ± 11	52.9	53.7	62.1	53.3
δ3′	128 ± 13	118.3	144.0	148.7	139.3
ε3′	184 ± 11	185.7	178.4	175.7	177.9

[a] M06-2X/6-31G(d,p) optimized geometries. All angles are defined in Figure 5.1. [b] Taken from Reference 37. [c] Average value is 258 ± 14° for purine and 241 ± 8° for pyrimidine nucleobases.

The calculated dG *anti*/*syn* energy differences in the natural deoxydinucleoside monophosphate sequences range from 1.9 to 9.6 kJ mol^{-1} depending on the dG position and the flanking base (Table 5.3). This range encompasses the *anti*/*syn* energy difference calculated using a nucleotide model (4.8 kJ mol^{-1}, Chapter 4). The energy differences are slightly larger when dG is in the 5' (6.4 – 9.6 kJ mol^{-1}) compared with the 3' (1.9 – 5.5

kJ mol^{-1}) position. This variation may arise due to differences in the conformation adopted by the deoxydinucleoside monophosphate to accommodate the *syn* conformation. Specifically, the relative base–base orientation is altered in *syn* 3′-dG sequences, which stabilizes the deoxydinucleoside monophosphate by maximizing favourable interactions between dG and the sugar-phosphate backbone while maintaining a backbone conformation similar to that for the corresponding *anti* dG structure (Figure 5.3a). In contrast, the backbone in the *syn* 5'-dG sequences distorts to accommodate the guanine ring over the sugar moiety without forming stabilizing backbone–base interactions, which maintains a nearly parallel relative base–base arrangement (Figure 5.3b). Most importantly, the relative energy calculated for all natural deoxydinucleoside monophosphates correctly predicts that dG preferentially adopts the *anti* conformation regardless of the position or flanking base. This suggests that neither distortion in the relative base–base orientation nor the sugar-phosphate backbone observed in this model is significant enough to stabilize the *syn* conformation of natural dG. The accurate prediction of the experimentally-observed structure and preferred *anti* conformation of natural dG further justifies the computational approach and indicates that the same methodology can be confidently applied to the C8-PhOH-dG damaged adducts.

Table 5.3 M06-2X relative energies (kJ mol^{-1}) of the *anti* and *syn* conformations of dG in deoxydinucleoside monophosphates [a]

Base	Position	Flanking base							
		Cytosine		Thymine		Adenine		Guanine	
		anti	*syn*	*anti*	*syn*	*anti*	*syn*	*anti*	*syn*
G	5′	0.0	9.6	0.0	7.4	0.0	6.4	0.0	8.6
	3′	0.0	1.9	0.0	5.5	0.0	2.0	0.0	2.2

[a] Relative energies calculated at the M06-2X/6-311+G(2df,p)//M06-2X/6-31G(d,p) level of theory and include scaled (0.9580) ZPVE corrections.

5.3.2 Deoxydinucleoside Monophosphates Containing C8-PhOH-dG Adducts

The orientations of the *anti o*- and *p*-PhOH-dG adduct nucleobases relative to the sugar moiety (χ, Table 5.4) fall within the range found in experimental crystal structures of natural B-DNA[38] regardless of the damage position (5′ versus 3′), and therefore are very close to those calculated for the corresponding natural sequences, as well as the isolated nucleotide adducts (245 – 250°, Chapter 4). In contrast, the χ values for the *syn* conformation of the adducts (82 – 97°, Tables A.2 and A.3, Appendix A) are slightly larger than those reported for *syn* dG (65 – 89°, Table A.1, Appendix A) for all sequences regardless of the damaged base (*ortho* versus *para*) or position. This agrees with the result for the isolated nucleotide adducts, which have χ angles (81.9 – 87.9°, Chapter 4) 14 – 20° greater than the *syn* dG nucleotide (67.9°, Chapter 4). Therefore, the calculated change in χ is likely an inherent property of the bulky adduct regardless of the environment. Nevertheless, it can be concluded that the bulky group does not significantly affect χ for the *anti* or *syn* conformation of natural dG.

Table 5.4 Average values of backbone torsional angles (degrees) in deoxydinucleoside monophosphates containing an *o*- or *p*-PhOH-dG adduct[a]

	o-PhOH-dG				*p*-PhOH-dG			
	anti		*syn*		*anti*		*syn*	
	5′	3′	5′	3′	5′	3′	5′	3′
χ5′	244.8	245.5	86.1	225.1	245.7	233.6	90.8	226.3
χ3′	255.0	252.1	260.5	90.9	250.7	267.3	248.6	87.9
γ5′	54.1	51.1	48.0	48.2	52.7	51.6	48.1	49.1
δ5′	85.95	149.1	139.7	142.4	99.0	148.6	140.9	143.8
ε5′	163.4	201.2	172.3	199.4	167.1	210.5	170.0	199.4
ζ	272.4	280.9	266.6	284.0	262.5	289.1	255.9	286.4
α	314.0	287.0	303.9	291.8	312.4	282.1	304.2	291.1
β3′	167.6	159.9	170.7	172.6	169.1	163.1	172.4	172.1
γ3′	59.2	55.5	66.3	50.3	58.4	52.1	63.0	51.5
δ3′	145.1	91.5	146.7	146.1	145.3	102.5	145.8	145.0
ε3′	177.0	193.9	176.4	175.8	177.1	190.1	176.2	175.6

[a] M06-2X/6-31G(d,p) optimized geometries. All angles are defined in Figure 5.1.

Figures 5.4 and 5.5 display examples of optimized deoxydinucleoside monophosphates containing the *o*-PhOH-dG and *p*-PhOH-dG phenoxyl adduct, respectively. The natural nucleobases in these sequences generally adopt the same orientation observed in the corresponding natural deoxydinucleoside monophosphate, and therefore are representative of the conformations in natural B-DNA. However, the orientation of the adduct nucleobase relative to the natural base is somewhat dependent upon the position (5′ versus 3′), conformation (*anti* versus *syn*) and type (*ortho* versus *para*) of damage. The unique features of each class of deoxydinucleoside monophosphate will be discussed separately below.

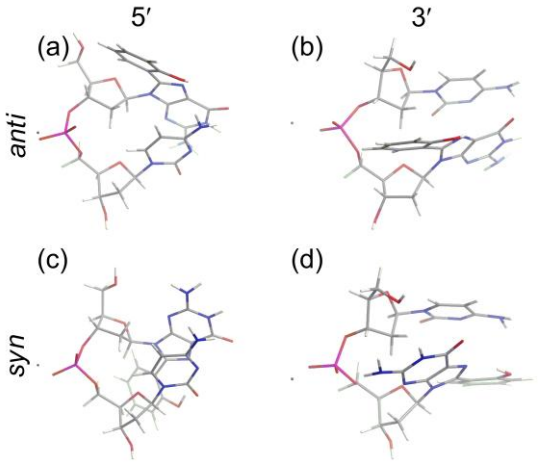


Figure 5.4 M06-2X/6-31G(d,p) structure of *o*-PhOH-dG in the 5′ (a, c) or 3′ (b, d) position and the *anti* (a, b) or *syn* (c, d) conformation, and C as the flanking base.

When *anti o*-PhOH-dG is located in the 5' position, the hydroxyl group on the phenoxyl ring always forms a (donating or accepting) hydrogen-bonding interaction with the 3' nucleobase (see, for example, Figure 5.6a). These interactions lead to a greater degree of tilt (φ) between the dG component of the adduct and the 3' nucleobase (by 2.5 – 12.6°, Table 5.1) compared with the corresponding natural strand. Additionally, these structures exhibit a greater degree of twist (θ) about the carbon-carbon bond connecting the phenoxyl ring and guanine base (by approximately 10 – 20°, Table 5.1) compared with

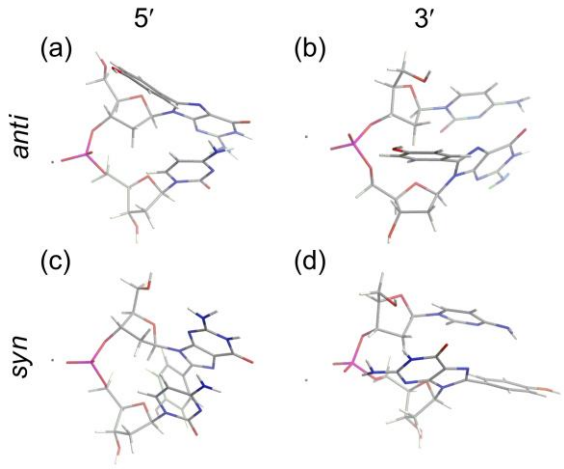


Figure 5.5 M06-2X/6-31G(d,p) structure of *p*-PhOH-dG in the 5′ (a, c) or 3′ (b, d) position and the *anti* (a, b) or *syn* (c, d) conformation, and C as the flanking base.

the isolated nucleotide adduct (θ = 159.7°, Chapter 4). Regardless of deviations in θ, the highly stabilizing O–H···N7 intramolecular hydrogen bond that was found in nucleoside (Chapter 2) and nucleotide (Chapter 4) models is still present when the adduct is incorporated into a deoxydinucleoside monophosphate complex. When the *p*–PhOH-dG adduct is considered in the same position, there are only small variations in θ compared to the damaged nucleotide (θ = 148.6°, Chapter 4) when any natural base is in the 3′ position (Table 5.1). Similar θ values coupled with the lack of a strong interaction between the bulky group and the 3′ base leads to base–base arrangements comparable to those discussed for the natural sequences involving dT, dC, or dA (Table 5.1). In contrast, there is a contact between the C–H_{phenyl} of 5′-*p*-PhOH-dG and O6 of 3′-dG (Figure 5.6b), which leads to a greater base–base tilt as discussed for sequences containing 5′-*o*-PhOH-dG.

The *anti* bulky adduct is subject to more steric crowding when placed in the 3' position. Specifically, close contacts (2.2 Å or less) exist between hydrogen atoms on the phenoxyl ring and those at C2' in the 5′ nucleotide for both adducts (Figure 5.7, left), which may significantly destabilize these configurations. Nevertheless, these arrangements are accompanied by weakly attractive interactions between hydrogen atoms on the phenoxyl

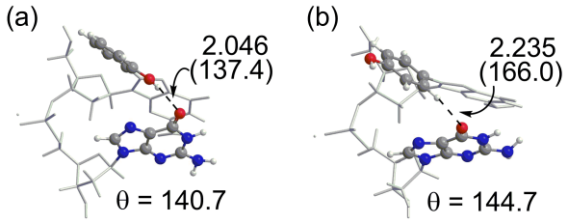


Figure 5.6 M06-2X/6-31G(d,p) structures [hydrogen-bond distances in Å (angles in degrees)] for the *anti* conformation of (a) *o*-PhOH-dG or (b) *p*-PhOH-dG in the 5′ position with G as the 3′-flanking base.

group and O5′ and/or a phosphate oxygen atom (Figure 5.7, right). The preferred θ in the *o*-PhOH-dG nucleotide adduct (159.7°, Chapter 4) is not affected by these interactions since the stabilizing O–H···N7 hydrogen bond anchors the relative arrangement of the phenoxyl and guanine rings (Figure 5.7a, right). However, since the O–H···N7 interaction is absent in *p*-PhOH-dG (Figure 5.7b, right), the steric contacts with the phenoxyl group significantly change θ (by up to 21°, Table 5.1) compared to the isolated nucleotide adduct (Chapter 4).

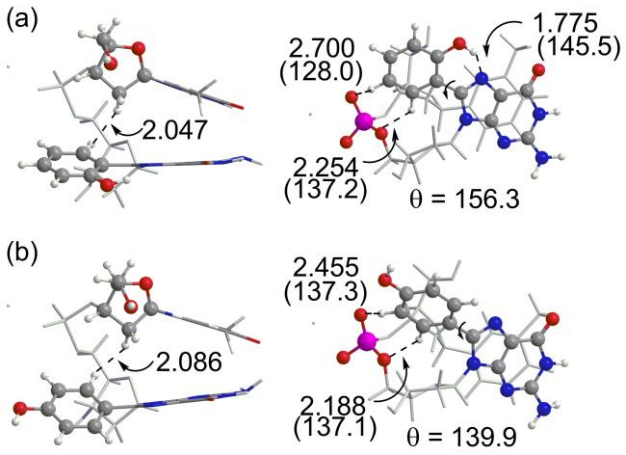


Figure 5.7 Edge (left) and face (right) views of M06-2X/6-31G(d,p) structures [hydrogen-bond distances in Å (angles in degrees)] for the *anti* conformation of (a) *o*-PhOH-dG or (b) *p*-PhOH-dG in the 3′ position with T as the 5′-flanking base.

Since the phenoxyl group of the *syn* conformation protrudes into the area corresponding to the minor groove of DNA, there is more steric hindrance when this conformation is adopted in the 5′ position of a deoxydinucleoside monophosphate

compared with the *anti* adduct. However, the *syn* conformation leads to O4' lone pair (lp)···π interactions with the phenoxyl ring (Figure 5.8a), which likely provide some stability.[39] As discussed for the 5' *anti* conformations, the *o*-PhOH-dG adduct exhibits more tilted relative base–base arrangements than the *p*-PhOH-dG adduct (Table 5.1). Interactions between the phenoxyl OH group in *o*-PhOH-dG and the 3' nucleobase (i.e., O4 carbonyl of dC or dT, N3 of dA, and N2–H in dG (Figure 5.8b)) lead to larger deviations (by 2 – 35°) from the preferred θ in the isolated nucleotide (θ = 203.4° *ortho*, θ = 217.1° *para*, Chapter 4).

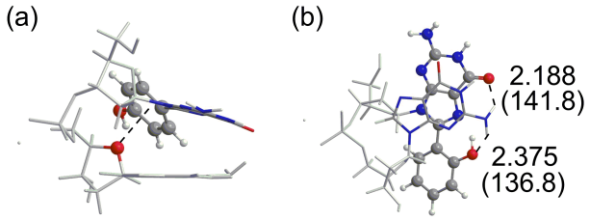


Figure 5.8 M06-2X/6-31G(d,p) structures [hydrogen-bond distances in Å (angles in degrees)] for the *syn* conformation of (a) *p*-PhOH-dG or (b) *o*-PhOH-dG in the 5′ position with G as the 3′-flanking base.

Compared with the 5′ position, there is less steric crowding between the phenoxyl group and neighboring bases when the *syn* adducts are in the 3' position, and therefore there are very small changes in the geometry of the adduct nucleotides compared with the isolated counterparts (Chapter 4). Specifically, θ changes by less than 10° for both adducts (Table 5.1). Most interestingly, the phenoxyl rings of the 3′ adducts can participate in stacking interactions with the 5′ nucleobase, where ring overlap is more predominant for 5' purines than pyrimidines (Figures 5.9 a and b). Since the dG and phenoxyl components of the adducts are twisted relative to one another, the participation of both adduct rings in stacking interactions with the neighbouring bases leads to significant tilt for these deoxydinucleoside monophosphate sequences in order to maximize ring–ring overlap (Table 5.1).

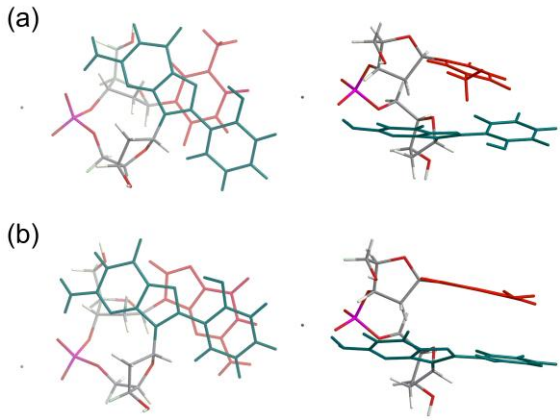


Figure 5.9 Face (left) and edge (right) views of the stacking interactions between the *syn* conformation of *o*-PhOH-dG (green) and the flanking base (red) when the adduct is in the 3′ position and the 5′ base is (a) a pyrimidine or (b) a purine.

The backbone angles for the *anti* conformation of the adducts averaged across all sequences are very similar to those found for natural dG (Tables 5.2 and 5.4). The largest changes occur in δ and ε when the adduct is in either (5′ or 3′) position since the sugar puckering changes from C2′–*endo* for natural dG (B-DNA) to C1′–*exo*-O4′–*endo* when the bulky group is positioned above the sugar. Although there are variations in the backbone configuration when the adducts adopt a *syn* versus *anti* conformation, there is less than a 15° distortion in a given backbone dihedral angle when either adduct adopts the *syn* conformation compared to the corresponding natural *syn* sequence (Table 5.2 and 5.4). Therefore, the majority of the distortion in the backbone upon adduct formation is likely due to the induced *syn* conformation of the base rather than the presence of the bulky C8 substituent.

The relative energy between the *anti* and *syn* conformations of the PhOH-dG adducts ranges between 0 and 21 kJ mol^{-1} (Table 5.5). There is no clear correlation between these energy differences and the type (*ortho* versus *para)* or position (5' versus 3') of the damage, or the nucleobase sequence. However, in contrast to natural dG, both

adducts exhibit a preference for the *syn* conformation. Indeed, the *syn* conformer is over 10 kJ mol^{-1} more stable than the corresponding *anti* conformer for the majority of the deoxydinucleoside monophosphates considered. The *syn* conformation of the bulky adducts may be stabilized to a greater extent than the *syn* conformation of natural dG partly due to the (hydrogen-bonding, stacking) interactions between the phenoxyl group and neighboring nucleotide discussed above. However, the primary reason for the change in the *anti/syn* conformational preference upon formation of bulky C8-phenoxyl adducts is likely destabilization of the *anti* conformation due to steric constraints arising from the sheer size of the bulky group and therefore close proximity to the sugar moiety.

Table 5.5 M06-2X relative energies (kJ mol^{-1}) of the *anti* and *syn* conformations of an *o*- or *p*-PhOH-dG adduct in deoxydinucleoside monophosphates[a,b]

Base	Position	Flanking base							
		Cytosine		Thymine		Adenine		Guanine	
		anti	*syn*	*anti*	*syn*	*anti*	*syn*	*anti*	*syn*
p–PhOH-G	5′	6.1	0.0	16.1	0.0	0.0	0.1	6.1	0.0
	3′	7.2	0.0	14.4	0.0	21.0	0.0	12.8	0.0
o–PhOH-G	5′	9.2	0.0	16.8	0.0	15.3	0.0	4.8	0.0
	3′	10.9	0.0	15.7	0.0	15.1	0.0	20.8	0.0

[a] Relative energies calculated at the M06-2X/6-311+G(2df,p)//M06-2X/6-31G(d,p) level of theory and include scaled (0.9580) ZPVE corrections. [b] Relative energies of the *syn* conformations compared to the *anti* conformations calculated using a nucleotide model are 4.8 kJ mol^{-1} for natural dG, 7.7 kJ mol^{-1} for the *ortho* adduct, and 1.5 kJ mol^{-1} for the *para* adduct (See Chapter 4).

The preferential stabilization of the *syn* conformation of the PhOH-dG adducts reported here contrasts the conclusion reached when smaller computational models are considered. Specifically, an isolated nucleotide model predicts the *anti* conformation to be more stable (by 1.5 (*para*) or 7.7 (*ortho*) kJ mol^{-1}, Chapter 4), which contrasts results for other (C8-methyl-dG[36,40] or 8-oxoguanine[41]) C8-substituted adducts, where a preference for the *syn* conformation is attributed to steric clashes between the C8-substituent and C2′-H.[36,41] However, the twist about θ may help relieve the steric clashes in the PhOH-dG nucleotide models, which explains why destabilization of the *anti* conformation is not

observed. When the deoxydinucleoside monophosphate model is considered, similar values of θ likely do not result in relief of steric strain, since the twist brings the bulky group of the adduct in the 3′ position into close proximity with the sugar of the 5′-flanking residue (Figure 5.7), while the 5′ adducts can undergo significant tilt and θ distortions to maximize interactions with the 3′-flanking base (Figure 5.6). Furthermore, the magnitude of the difference in relative energies is much larger for the deoxydinucleoside monophosphate systems than calculated using the smaller model. Since deoxydinucleoside monophosphate models are the most relevant to DNA thus far in this thesis, this emphasizes the importance of using larger computational models when assessing the *anti*/*syn* preferences of bulky DNA adducts, where steric clashes, as well as other (hydrogen-bonding, stacking) interactions, with neighboring bases may dictate the observed conformation.

Since the backbone orientation is proposed to play a role in DNA damage recognition and repair,[14] the optimized structures have implications for whether phenoxyl adducts are likely to be recognized by repair enzymes, or persist in the helix and potentially lead to mutations.[14] Specifically, for other bulky adducts, it has been hypothesized that structural distortions to the helix caused by the *syn* conformation may lead to recognition and repair,[14,42] while adoption of the *anti* conformation may permit the lesion to go unrecognized (until *anti/syn* conversion occurs).[43] The calculations in this chapter suggest that the backbone does not exhibit significant distortion upon adduct formation regardless of the (*anti/syn*) conformation adopted, and thus it is possible that the adduct will not be repaired.

Since the model system implemented in the present work represents the structure of natural DNA duplexes prior to replication,[9,10] the finding that the *syn* conformation of the adduct is consistently preferred, coupled with the above hypothesis that the adduct may not be repaired, has significant implications on the projected mutagenicity of phenoxyl

adducts. For example, as discussed in Chapter 1, AAF-dG adducts are known to adopt the *syn* conformation,[14,44] which largely results in frameshift mutations,[14,45] while AF-dG adducts can adopt either the *anti* or *syn* conformation[46] and primarily result in either base-substitution mutations or normal replication.[45,47] Therefore, if the phenoxyl adducts primarily adopt the *syn* conformation, base-substitution mutations or frameshift mutations may arise upon replication.[48] However, since the relatively small size of the phenoxyl group may not lead to bulge stabilization as seen for AAF-dG, another base may be readily accommodated opposite the adduct and base-substitution mutations may be favored over frameshift mutations.[14]

To confirm the biological implications of the C8 adduct based on the preferred structures found in this work, the following chapter must further extend the computational model to DNA duplexes. Indeed, even the structure of the natural deoxydinucleoside monophosphate depends on the location of *syn* dG, which further suggests both intrastrand neighboring bases could be important for elucidating the adduct structure and *anti/syn* energy difference in DNA. Furthermore, differences in the intramolecular contacts when the adducts are located in the 5' and 3' positions suggest that nucleotides on both sides of the damaged adduct must be simultaneously taken into account. Despite several explicit interactions between the phenoxyl group and the flanking base, the present chapter finds a lack of sequence dependence on the preferred adduct conformation. Since this contrasts findings for other bulky adducts,[1,5,49] sequence effects should be further explored in larger DNA duplexes. In particular, DNA duplexes should contain pyrimidine flanking bases simultaneously in the 5′ and 3′ positions, as well as purines simultaneously in both positions, since these two environments represent the extremes in base–base overlap observed in the deoxydinucleoside monophosphate model. While the current study suggests intrastrand base–base and backbone interactions lead to a preference for the *syn*

conformation, Chapter 3 found that complementary base-pairing with the adducts is stronger in the *anti* conformation. In addition, discrete interactions between the phenoxyl moiety and flanking bases may affect the hydrogen-bonding patterns[50] of the adduct and therefore its base-pairing preference. Therefore, simulations of DNA duplexes with the adducts paired with the strongest (cytosine and guanine, Chapter 3) binding partners for each (*anti* and *syn*) conformation are necessary to determine the most likely conformations and hydrogen-bonding patterns adopted by the adducts in biological systems. These models will allow the combined effects of stacking and hydrogen bonding on the conformational and base-pairing preferences to be studied.

5.4 Conclusions

The dG component of the bulky *o*- and *p*-PhOH-dG adducts assumes a similar relative base–base orientation and backbone conformation as the corresponding natural dG in a deoxydinucleoside monophosphate model. However, the conformation of the phenoxyl group can be characterized according to the position (5′ versus 3′), conformation (*anti* versus *syn*) and type (*ortho* versus *para*) of damage. The *ortho* adduct has greater conformational flexibility since the hydroxyl group can interact with neighboring bases and thereby cause distortion to the local structure. In particular, the twist angle between the phenoxyl group and G base (θ) is highly affected by these interactions, which in turn may alter the tilt angle (φ) between the neighboring bases or distort the backbone to improve ring–ring overlap between the bases. In contrast, the *para* adduct is relatively unaffected by these considerations.

The most striking conclusion from this chapter is the change in relative energies of the *anti* and *syn* conformations upon DNA damage. In particular, when the adduct is in the 3′ position, the *anti* conformation is destabilized by steric clashes between the phenoxyl group and the 5′ sugar, as well as the phosphate backbone. Although there is often a high

degree of tilt when the adduct is in the 5′ position to maintain stabilizing interactions with the 3′-flanking base, these interactions are insufficient to overcome destabilization due to the steric clash between the bulky group and the sugar. In the *syn* conformation, overlap between the adduct and the flanking base is enhanced with respect to *syn* dG in the natural sequences due to the involvement of the phenoxyl ring, especially when the adduct is in the 3′ position. In the 5′ position, the phenoxyl ring of the adduct forms lone pair (lp)···π interactions with O4′ of the 3′-flanking base. Therefore, while a clear preference for the *anti* conformation exists in the unmodified sequences, addition of the bulky phenoxyl group results in a preference for the *syn* conformation for both adducts regardless of sequence or position. This result was not observed for smaller models (nucleoside, nucleotide) in previous chapters, which stresses the importance of using models that address both base–base and backbone interactions when determining the conformational preferences of bulky adducts. In addition, the magnitude of the energy difference is much more significant for the deoxydinucleoside monophosphate model than for the nucleotide model in Chapter 4.

Since replication errors for other C8 bulky adducts have been proposed to result from a preference for the *syn* conformation, these findings provide insight into the possible mutagenicity of phenoxyl adducts. To further clarify the preferred adduct conformation in DNA, the model system must be expanded to simultaneously include both 5′ and 3′ flanking bases, which may further increase the *syn* preference. In addition, hydrogen-bonding and steric interactions with the complementary strand should considered to determine whether the preferred conformation is dependent on the base pairing of the adduct, as well as whether mismatches are stabilized. Although the lack of sequence dependence found in this model suggests the *syn* conformation may be adopted regardless of the identity of flanking bases or (*o*- or *p*-PhOH-dG) adduct type, discrete interactions between the adduct and flanking bases may be important in determining the orientation of the bulky group for

each adduct in DNA. This could subsequently impact base pairing since results in Chapter 3 show that the phenoxyl group can participate in hydrogen bonding. MD simulations will be explored in the following chapter to incorporate all of the above structural influences into a single system that will increase our understanding of the conformational and base-pairing properties, as well as biological implications, of phenoxyl DNA damage.

5.5 References

(1) Jain, N.; Meneni, S.; Jain, V.; Cho, B. P. *Nucleic Acids Res.* **2009**, *37*, 1628-1637.

(2) Cho, B. P.; Beland, F. A.; Marques, M. M. *Biochemistry* **1994**, *33*, 1373-1384.

(3) Eckel, L. M.; Krugh, T. R. *Nat. Struct. Biol.* **1994**, *1*, 89-94.

(4) Eckel, L. M.; Krugh, T. R. *Biochemistry* **1994**, *33*, 13611-13624.

(5) Elmquist, C. E.; Wang, F.; Stover, J. S.; Stone, M. P.; Rizzo, C. J. *Chem. Res. Toxicol.* **2007**, *20*, 445-454.

(6) Poltev, V. I.; Anisimov, V. M.; Danilov, V. I.; Deriabina, A.; Gonzalez, E.; Jurkiewiez, A.; Les, A.; Polteva, N. *J. Biomol. Struct. Dyn.* **2008**, *25*, 563-571.

(7) Poltev, V. I.; Anisimov, V. M.; Danilov, V. I.; Deriabina, A.; Gonzalez, E.; Garcia, D.; Rivas, F.; Jurkiewicz, A.; Les, A.; Polteva, N. *Journal of Molecular Structure: THEOCHEM* **2009**, *912*, 53-59.

(8) Poltev, V. I.; Anisimov, V. M.; Danilov, V. I.; Mourik, T. v.; Deriabina, A.; González, E.; Padua, M.; Garcia, D.; Rivas, F.; Polteva, N. *Int. J. Quantum Chem.* **2010**, *110*, 2548–2559.

(9) Shapiro, R.; Sidawi, D.; Miao, Y. S.; Hingerty, B. E.; Schmidt, K. E.; Moskowitz, J.; Broyde, S. *Chem. Res. Toxicol.* **1994**, *7*, 239-253.

(10) Hingerty, B.; Broyde, S. *Biochemistry* **1982**, *21*, 3243-3252.

(11) Meneni, S. R.; Shell, S. M.; Gao, L.; Jurecka, P.; Lee, W.; Sponer, J.; Zou, Y.; Chiarelli, M. P.; Cho, B. P. *Biochemistry* **2007**, *46*, 11263-11278.

(12) Meneni, S.; Shell, S. M.; Zou, Y.; Cho, B. P. *Chem. Res. Toxicol.* **2007**, *20*, 6-10.

(13) Mao, B.; Hingerty, B. E.; Broyde, S.; Patel, D. J. *Biochemistry* **1998**, *37*, 95-106.

(14) Patel, D. J.; Mao, B.; Gu, Z. T.; Hingerty, B. E.; Gorin, A.; Basu, A. K.; Broyde, S. *Chem. Res. Toxicol.* **1998**, *11*, 391-407.

(15) Gu, Z. T.; Gorin, A.; Hingerty, B. E.; Broyde, S.; Patel, D. J. *Biochemistry* **1999**, *38*, 10855-10870.

(16) McLaughlin, C. K.; Lantero, D. R.; Manderville, R. A. *J. Phys. Chem. A* **2006**, *110*, 6224-6230.

(17) Tereshko, V.; Minasov, G.; Egli, M. *J. Am. Chem. Soc.* **1999**, *121*, 470-471.

(18) Gorb, L.; Shishkin, O.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2005**, *22*, 441-454.

(19) Kosenkov, D.; Gorb, L.; Shishkin, O. V.; Sponer, J.; Leszczynski, J. *J. Phys. Chem. B* **2008**, *112*, 150-157.

(20) Kosenkov, D.; Kholod, Y. A.; Gorb, L.; Shishkin, O. V.; Kuramshina, G. M.; Dovbeshko, G. I.; Leszczynski, J. *J. Phys. Chem. A* **2009**, *113*, 9386-9395.

(21) Leulliot, N.; Ghomi, M.; Scalmani, G.; Berthier, G. *J. Phys. Chem. A* **1999**, *103*, 8716-8724.

(22) Palamarchuk, G. V.; Shishkin, O. V.; Gorb, L.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2009**, *26*, 653-661.

(23) Rubio, M.; Roca-Sanjuan, D.; Serrano-Andres, L.; Merchan, M. *J. Phys. Chem. B* **2009**, *113*, 2451-2457.

(24) Shishkin, O. V.; Gorb, L.; Zhikol, O. A.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2004**, *21*, 537-553.

(25) Shishkin, O. V.; Gorb, L.; Zhikol, O. A.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2004**, *22*, 227-243.

(26) Shishkin, O. V.; Palamarchuk, G. V.; Gorb, L.; Leszczynski, J. *J. Phys. Chem. B* **2006**, *110*, 4413-4422.

(27) Sponer, J.; Sabat, M.; Gorb, L.; Leszczynski, J.; Lippert, B.; Hobza, P. *J. Phys. Chem. B* **2000**, *104*, 7535-7544.

(28) Svozil, D.; S , A.; Cheatham III, T. E.; Forti, F.; Luque, F. J.; Orozco, M.; Sponer, J. i. *J. Phys. Chem. B* **2008**, *112*, 8188-8197.

(29) Zakjevskii, V. V.; Dolgounitcheva, O.; Zakrzewski, V. G.; Ortiz, J. V. *Int. J. Quantum Chem.* **2007**, *107*, 2266-2273.

(30) Scalmani, G.; Frisch, M. J. *J. Chem. Phys.* **2010**, *132*, 114110(15).

(31) Dabkowska, I.; Gonzalez, H. V.; Jurecka, P.; Hobza, P. *J. Phys. Chem. A* **2005**, *109*, 1131-1136.

(32) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215-241.

(33) Merrick, J. P.; Moran, D.; Radom, L. *J. Phys. Chem. A* **2007**, *111*, 11683-11700.

(34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Jr., J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; Revision A.02 ed.; Gaussian, Inc.: Wallingford CT, 2009.

(35) Heavner, S.; Gannett, P. M. *J. Biomol. Struct. Dyn.* **2005**, *23*, 203-219.

(36) Vongsutilers, V.; Phillips, D. J.; Train, B. C.; McKelvey, G. R.; Thomsen, N. M.; Shaughnessy, K. H.; Lewis, J. P.; Gannett, P. M. *Biophys. Chem.* **2011**, *154*, 41-48.

(37) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag New York Inc.: New York, NY, 1984.

(38) Schneider, B.; Neidle, S.; Berman, H. M. *Biopolymers* **1997**, *42*, 113-124.

(39) Egli, M.; Sarkhel, S. *Accounts Chem. Res.* **2007**, *40*, 197-205.

(40) Kohda, K.; Tsunomoto, H.; Minoura, Y.; Tanabe, K.; Shibutani, S. *Chem. Res. Toxicol.* **1996**, *9*, 1278-1284.

(41) Gannett, P. M.; Sura, T. P. *Chem. Res. Toxicol.* **1993**, *6*, 690-700.

(42) Geacintov, N. E.; Broyde, S.; Buterin, T.; Naegeli, H.; Wu, M.; Yan, S. X.; Patel, D. J. *Biopolymers* **2002**, *65*, 202-210.

(43) Perlow, R. A.; Broyde, S. *J. Mol. Biol.* **2002**, *322*, 291-309.

(44) Wang, L. H.; Broyde, S. *Nucleic Acids Res.* **2006**, *34*, 785-795.

(45) Bichara, M.; Fuchs, R. P. P. *J. Mol. Biol.* **1985**, *183*, 341-351.

(46) Mao, B.; Hingerty, B. E.; Broyde, S.; Patel, D. J. *Biochemistry* **1998**, *37*, 81-94.

(47) Cho, B. S. P. *J. Environ. Sci. Health Pt. C-Environ. Carcinog. Ecotoxicol. Rev.* **2004**, *22*, 57-90.

(48) Liang, F. T.; Cho, B. P. *Chem. Res. Toxicol.* **2011**, *24*, 597-605.

(49) Wang, F.; Elmquist, C. E.; Stover, J. S.; Rizzo, C. J.; Stone, M. P. *Biochemistry* **2007**, *46*, 8498-8516.

(50) Dupradeau, F. Y.; Case, D. A.; Yu, C. Z.; Jimenez, R.; Romesberg, F. E. *J. Am. Chem. Soc.* **2005**, *127*, 15612-15617.

6 Chapter 6: Molecular Dynamics Simulations of DNA Containing C-Linked PhOH-dG Adducts[a]

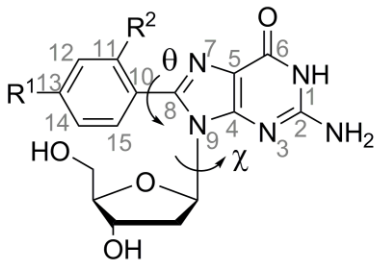
6.1 Introduction

The structure of the C-linked PhOH-dG adducts has been examined in depth in previous chapters from a small model perspective, which allows the individual effects of various DNA environmental factors (hydrogen bonding, backbone, flanking bases) to be investigated. This chapter simultaneously examines the conformational and base-pairing properties of the adducts using a large model approach. Specifically, this chapter expands the current understanding of the C-linked phenoxyl adducts by considering the structural impact of *p*-PhOH-dG and *o*-PhOH-dG within the two decanucleotides (ODN1 and ODN2) shown in Figure 6.1. In ODN1, the C-linked adducts (X) are flanked by pyrimidine bases, while in ODN2 they are flanked by purines. This allows for comparison of the two extremes of flanking base environments, where Chapter 5 found that base–base overlap differs significantly when purines or pyrimidines are in the flanking site. In addition, recent experimental studies[1] determined melting temperatures using these sequences to determine relative duplex stability. However, the experimental results were unable to conclusively determine the conformational preferences of the adducts, and therefore this work was done in collaboration, where the combined experimental and computational approach leads to a deeper understanding of the results than provided by either technique alone.

In Chapter 3, the effect of hydrogen-bonding interactions on the *anti*/*syn* conformational preference showed that *o*-PhOH-dG forms significantly more stable complexes with the Watson-Crick face than with the Hoogsteen face, and that C forms the

---

[a] Reprinted in part with permission from Chem. Res. Toxicol., submitted for publication. Omumi, A.; Millen, A. L.; Wetmore, S. D.; Manderville, R. A. Unpublished work copyright 2011 American Chemical



a: R^1 = OH, R^2 = H = *para* (*p*)-PhOH-dG
b: R^1 = H, R^2 = OH = *ortho* (*o*)-PhOH-dG

ODN1 = 5′-CCATXCTACC, X= a or b
ODN1′(N) = 5′-GGTAGNATGG, N = C or G
ODN2 = 5′-GGTAGXATGG, X = a or b
ODN2′(N) = 5′-CCATNCTACC, N = C or G

Figure 6.1 Oligonucleotide sequences considered in this chapter.

most stable base pair with *o*-PhOH-dG regardless of the hydrogen-bonding face involved. However, Hoogsteen interactions are stronger for *p*-PhOH-dG than those with natural dG and the *para* adduct binds with nearly equal stability to both C and G in the Hoogsteen orientation. Therefore, in the complementary strands, ODN1′(N) and ODN2′(N), the base opposite (N) the adduct is varied (N = C or G) based on the conclusions from Chapter 3 to determine the possibility of mismatch stabilization by the modified C-linked nucleobases. This work represents an important step in understanding the mutagenic profile of C-linked PhOH-dG adducts by combining the significant effects of backbone, flanking bases, and complementary bases, and thereby conclusively determining the adduct conformational and base-pairing preferences in DNA duplex environments.

6.2 Computational Details

In molecular dynamics (MD) simulations of the oligonucleotides, parameters for the adducts were taken from the parmbsc0 force field,[2] since this is the most recently developed force field for nucleic acids, with the exception of dihedral parameters for the

---

Society. Manuscript ID: tx-2011-00247f.

phenoxyl moiety which were adopted from previous literature (parm94 force field).[3] ANTECHAMBER 1.4 was used to assign atom types.[4] The R.E.D.v.III.4 program[5] was used to calculate the partial charges of the modified bases using a multiconformation and multiorientation approach,[5] as well as the rigid-body re-orientation algorithm.[6] Two input structures corresponding to the lowest energy *anti* and lowest energy *syn* structures previously identified in Chapter 4 were used in the multiconformation approach. Intermolecular charge constraints between the nucleoside sugar moiety and a dimethylphosphate group were used, as well as intermolecular charge equivalencing procedures,[5] to obtain partial charges for the sugar moiety in the adducts analogous to the four natural nucleosides in the parm99 force field.[7] Partial charges were obtained from a two stage RESP fitting and Hartree-Fock single-point calculations with the 6-31G(d) basis set using the R.E.D. code interfaced with Gaussian 03[8] and ANTECHAMBER 1.4.[4]

Simulations were carried out using the SANDER module of the AMBER 10 or 11 software packages.[9,10] The parmbsc0 modification[2] to the parm99 force field[7] was used for the simulations. Initial structures were prepared using the NAB program[9] and GaussView[11] to modify the C8-PhOH-dG residue. The LEaP module of AMBER 10 was used to prepare the systems for MD. The (natural or modified) DNA strand was neutralized with 18 sodium ions and solvated with an 8 Å TIP3P octahedron water box.[12] Coulombic interactions were approximated using the particle-mesh Ewald method,[13] bonds involving hydrogen atoms were constrained using the SHAKE option,[14] and a 2 fs timestep was used throughout the simulation. A 10 Å cut-off was applied to Lennard-Jones interactions, and periodic boundary conditions were implemented. 500 steps of steepest descent minimization followed by 500 steps of conjugate gradient minimization were performed with the solute held fixed using a positional restraint with a force constant of 500 kcal mol^{-1} $Å^{-2}$. Subsequently, 1000 steps of steepest descent minimization followed by 1500 steps of

conjugate gradient minimization were performed on the entire system. The system was heated from 0 to 300 K using the Langevin thermostat with the solute restrained using a force constant of 10 kcal mol^{-1} $Å^{-2}$ over 20 ps under constant volume conditions. Following heating, a total of 20 ns of unrestrained MD simulations were carried out on each system under constant temperature (300 K) and pressure (1 atm) conditions. Snapshots were taken every 500 fs, and average structures were calculated from the final 2 ns of the simulation.

Percent occupancy throughout the trajectory was calculated for important hydrogen bonds with less than 3 Å heavy atom separation and 120° X–H–X angle. Binding interactions were calculated for MD structures obtained by averaging the coordinates of the frames from the final 2ns of the trajectory, and replacing the backbone and surrounding bases with an H atom. ΔE_{HBond} is calculated using counterpoise-corrected B3LYP/6-311+G(2df,p) energies. Stacking interactions were similarly calculated for MD structures obtained by averaging the coordinates of select frames with θ values representative of the dominant conformation of the adduct (shown in Figures 6.3 and 6.4), and replacing the backbone and surrounding bases with an H atom. Stacking binding strengths were calculated with M06-2X/6-31+G(d,p) since this combination has been shown to produce reliable interaction energies for noncovalent interactions that have large dispersion components.[15] $\Delta E_{Int5'}$ is defined as the interaction energy between X and the 5′-flanking base (G for ODN2 or T for ODN1), while $\Delta E_{Int3'}$ is the interaction energy between X and the 3′-flanking base (A for ODN2 or C for ODN1). All interaction energies are calculated as the energy difference between the dimer and the individual monomers in the dimer geometry. All DFT calculations were performed with Gaussian 03[8] or Gaussian 09.[16]

6.3 Results

MD simulations on the unmodified ODN1:1′(C) and ODN2:2′(C) sequences (X = G), yield DNA duplexes with average B-DNA backbone dihedral angles (Table 6.1) and three highly-occupied (70 – 89%) Watson-Crick hydrogen-bonds between X and C. The average structure of the X:C base pair has a stability of –106.7 kJ/mol for ODN1:1′ and –108 kJ/mol for ODN2:2′ (Table 6.2). When X = *anti* G is mispaired with *syn* G, two strong G:G (~50 kJ/mol) Hoogsteen bonds form with 35 and 66% occupancy in ODN1:1′ and 67 and 55% occupancy in ODN2:2′. The strands containing the mispair are likely less stable than native DNA helices due to weaker G:N hydrogen-bond strengths (by approximately 50 kJ/mol, Table 6.3). The strength of stacking interactions between X = G and the 5′-flanking base is –12 kJ/mol in ODN1:1′(C) and –6 kJ/mol in ODN2:2′(C), while the interactions with the 3′-flanking base are –50 kJ/mol in ODN1:1′(C) and –25 kJ/mol in ODN2:2′(C). For G:G mismatches, the stacking interactions range between –13 and –27 kJ/mol for the 5′-flanking base, and –6 and –42 kJ/mol for the 3′-flanking base.

When X is modified to the *o*-PhOH- or *p*-PhOH-dG adduct, there are no large structural changes to the double helix. The strands retain average values of B-DNA backbone dihedral angles with the exception of some local distortion at the site of the adduct (Table 6.1). All natural and modified bases remain within the helix regardless of the (*anti* or *syn*) conformation adopted (Figure 6.2). Thus, although intercalation of related bulky adducts that adopt the *syn* conformation into the helix can displace the opposing base,[17,18] the relatively small size of the C8-phenoxyl group is well accommodated opposite another base within the helix. However, the structure of the adduct nucleotide can significantly deviate from that predicted by the smaller (isolated nucleotide) computational model (Chapter 4). In particular, the dihedral angle θ can vary widely, depending on the adduct, the sequence, and the (*anti*/*syn*) conformation. Therefore, further details of the

Table 6.1 Average values of backbone dihedral angles (degrees) according to snapshots from MD simulations

N	Conf	ODN	X		α	β	γ	δ	ε	ζ	χ
–	–	– [a]	–		298±15	176±9	48±11	128±13	184±11	265±10	– [b]
C	*anti*	1	G	Avg[c]	288.9	168.9	61.2	121.6	199.4	255.5	241.0
				X[d]	283.7	172.2	63.0	134.0	203.4	232.6	249.1
			b	Avg[c]	274.5	173.7	71.3	126.9	195.0	259.6	231.2
				X[d]	280.2	174.6	45.5	127.4	282.4	127.2	286.2
			a	Avg[c]	288.6	171.1	58.5	118.0	189.9	269.5	234.9
				X[d]	301.4	167.4	57.0	87.6	180.4	269.8	227.8
		2	G	Avg[c]	286.2	164.7	67.5	117.7	194.4	261.6	242.9
				X[d]	288.3	163.6	56.2	110.5	214.9	236.8	228.2
			b	Avg[c]	291.3	170.8	59.3	116.0	189.8	268.0	234.5
				X[d]	299.3	163.6	62.4	78.6	181.3	283.1	215.1
			a	Avg[c]	290.1	174.7	56.3	121.0	186.6	266.5	236.4
				X[d]	311.8	168.7	55.9	81.4	189.5	277.9	217.3
	syn	1	b	Avg[c]	287.4	173.1	64.0	121.6	204.2	261.7	218.3
				X[d]	294.0	172.2	52.2	128.0	177.6	283.4	74.1
			a	Avg[c]	282.5	173.2	56.4	114.3	190.4	262.8	236.6
				X[d]	286.8	169.5	57.4	116.2	188.2	269.4	65.2
		2	b	Avg[c]	285.5	170.1	66.3	118.5	200.4	253.7	244.5
				X[d]	296.4	175.4	56.4	130.4	181.9	282.8	67.7
			a	Avg[c]	288.1	162.7	69.6	117.5	195.9	262.4	232.3
				X[d]	286.8	175.2	60.7	115.4	195.0	282.2	61.1
G	*anti*	1	G	Avg[c]	285.6	171.3	63.2	116.7	188.0	272.0	237.2
				X[d]	286.6	172.8	52.9	140.2	251.0	175.9	261.6
			b	Avg[c]	285.4	170.7	68.1	113.7	198.8	275.1	235.9
				X[d]	270.3	66.0	180.4	110.6	193.4	274.4	213.9
			a	Avg[c]	278.0	173.8	66.2	119.9	192.0	266.3	237.8
				X[d]	278.4	176.5	53.7	123.1	260.3	152.4	270.5
		2	G	Avg[c]	286.0	171.7	73.7	116.1	192.6	269.9	238.3
				X[d]	273.1	114.2	115.7	123.8	196.2	283.4	221.8
			b	Avg[c]	283.7	171.0	56.8	124.5	199.8	256.9	243.0
				X[d]	291.0	170.9	49.0	116.2	269.8	150.0	270.0
			a	Avg[c]	280.7	171.1	68.4	116.8	198.5	263.0	239.8
				X[d]	235.2	90.2	185.1	111.0	195.5	276.6	214.6
	syn	1	G	Avg[c]	289.3	165.0	62.0	115.6	193.2	268.0	219.3
				X[d]	294.8	165.5	57.8	102.7	199.4	265.1	55.9
			b	Avg[c]	293.2	168.0	56.9	109.4	191.1	268.7	227.9
				X[d]	292.1	170.7	55.6	122.4	183.4	281.5	72.1
			a	Avg[c]	291.0	170.4	62.4	108.7	192.2	271.6	234.1
				X[d]	294.2	168.1	54.3	118.9	177.1	282.8	70.0
		2	G	Avg[c]	289.9	170.2	61.4	120.2	200.5	253.0	242.0
				X[d]	293.2	169.8	54.0	109.3	190.5	271.7	59.6
			b	Avg[c]	288.4	172.6	64.4	116.4	191.2	266.2	237.6
				X[d]	293.8	170.1	49.8	110.6	179.4	275.8	73.7
			a	Avg[c]	284.4	166.8	57.2	113.6	201.1	249.8	229.7
				X[d]	296.8	177.6	61.8	134.9	192.9	280.9	65.0

[a] Average values from crystal structures of B-DNA.[19] [b] Average value for purine nucleobases is 258 ± 14° and 241 ± 8° for pyrimidines. [c] Averages were calculated using the Curves+ program and snapshots at 6, 10, 14, and 18 ns and backbones for every nucleotide except X.

[d] Average values for X were calculated using the Curves+ program and snapshots at 6, 10, 14, and 18 ns.

structure of each adduct in each sequence when paired with C or mismatched with G will be discussed below, where results are compared to the isolated nucleotide adduct since this is the most relevant model to DNA that is not specific to a particular sequence. Hydrogen-bond stabilities and occupancies (percentage of the total simulation time) will also be presented, where greater occupancies represent more stable hydrogen bonds. In addition, the strength of (intrastrand) stacking interactions will be discussed to aid determination of sequence effects on adduct structure and stability.

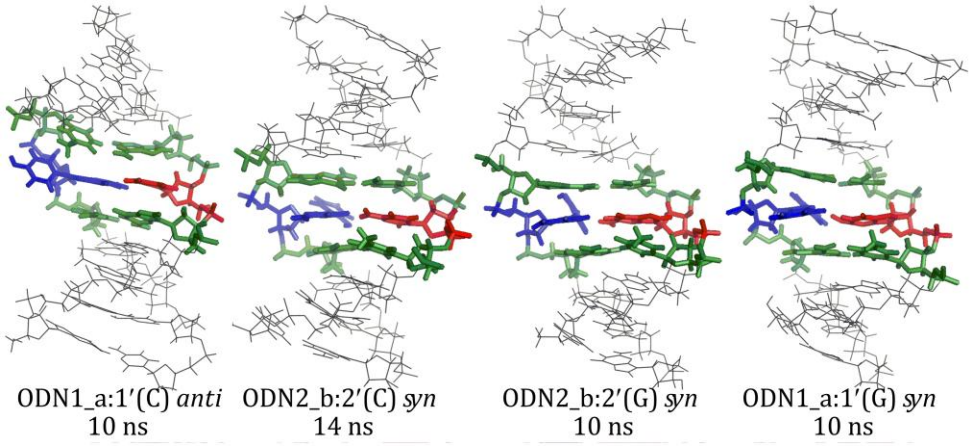


Figure 6.2 Examples of MD snapshots showing the view into the major groove of the B-DNA structure of the modified ODN1 and ODN2 duplexes with X = *p*-PhOH-dG (a) or *o*-PhOH-dG (b) (blue) paired opposite N = C or G (red) (purine or pyrimidine flanking bases highlighted in green).

6.3.1 *Anti* Conformation of Adducts (X = a or b) for N = C:

When the phenoxyl adducts (a and b) adopt the *anti* conformation in ODN1:1′(C) and ODN2:2′(C), the bulky phenoxyl group is located in the major groove. Three strong Watson-Crick hydrogen bonds are intact (~ –100 kJ/mol, Table 6.2) between X:C (66 – 84% occupancy, Tables 6.3 and 6.4). Stacking between X and the flanking bases is generally

stronger compared to the unmodified sequences (by up to 14 kJ/mol, Table 6.2). However, the adduct structure is sequence dependent. For ODN2:2′(C), the average θ of both adducts throughout the simulation (221.2 – 224.7°, Figures 6.3 and 6.4) is similar to the minimum identified using smaller computational models, which predict a small θ twist (by 20°, Chapter 4) for *o*-PhOH-dG due to a strongly stabilizing (O–H···N7) intramolecular hydrogen bond, and only a slightly greater twist (by 31°, Chapter 4) for *p*-PhOH-dG. The hydroxyl group of *o*-PhOH-dG remains hydrogen bonded to N7 for 73% of the simulation (Table 6.4). For ODN1_b:1′(C), the θ angle in *o*-PhOH-dG is highly twisted (θ = 98.9°) throughout the majority of the simulation (Figure 6.3), which corresponds to a structure where the OH group is directed away from the dG nucleobase. Correspondingly, the O–H···N7 hydrogen bond is disrupted (only 6% occupancy, Table 6.4). Instead, there is a very strong (–60 kJ/mol, Table 6.2) interaction between the phenoxyl oxygen in the adduct and the flanking 3′-dC amino group for 19% of the simulation (Table 6.4). In ODN1_a:1′(C), *p*-PhOH-dG has a different average θ (127.7°), which results from three minimum conformations (θ = ∼40°, ∼220°, and ∼310°, respectively, Figure 6.4), than *o*-PhOH-dG since the location of the hydroxyl group does not afford the same hydrogen bonds. The difference in θ, and the corresponding difference in hydrogen-bonding patterns for *o*-PhOH-dG, within ODN1 versus ODN2 suggests that the adduct structure is dependent on the sequence. Regardless, the similarity in hydrogen bonding and stacking between strands containing the *anti* conformation of the adducts and natural dG may result in similar stability between the natural and modified strands.

6.3.2 *Syn* Conformation of the Adducts (X = a or b) for N = C

When *o*-PhOH-dG adopts the *syn* conformation within ODN1_b:1′(C) and ODN2_b:2′(C), there is much greater distortion in the base pairs compared with the corresponding *anti* pairs. This is due to the decreased stability of X:C Hoogsteen hydrogen

Table 6.2 DFT Interaction Energies in Average Conformations Obtained from MD Simulations

Conformation	X	N	Strand	ΔE_{HBond} [a]	$\Delta E_{Int5'}$ [b]	$\Delta E_{Int3'}$ [c]
anti	G	C	ODN1	–106.7	–12.2	–50.1
			ODN2	–108.7	–6.0	–24.8
		G	ODN1	–49.9	–22.5	–42.9
			ODN2	–50.5	–13.5	–34.4
	a	C	ODN1	–105.6	–17.3	–51.6
			ODN2	–106.4	–18.6	–37.7
		G	ODN1	–22.7	–15.5	–46.3
			ODN2	–49.0	–12.6	–33.6
	b	C	ODN1	–105.3	–10.1	–59.7
			ODN2	–107.1	–14.9	–39.1
		G	ODN1	–26.3	–19.7	–54.2
			ODN2	–55.7	–13.0	–33.7
syn	G	C	ODN1	N/A	N/A	N/A
			ODN2	N/A	N/A	N/A
		G	ODN1	–50.5	–13.8	–6.3
			ODN2	–49.9	–27.9	–13.4
	a	C	ODN1	–35.5	–16.6	–7.0
			ODN2	–3.6	–55.8	–18.2
		G	ODN1	–46.9	–16.5	–14.2
			ODN2	–48.2	–35.5	–19.9
	b	C	ODN1	6.0	–13.5	–6.6
			ODN2	–24.7	–36.5	–27.3
		G	ODN1	–45.5	–16.8	–7.9
			ODN2	–41.3	–44.0	–20.4

[a] ΔE_{HBond} is calculated as the counterpoise-corrected B3LYP/6-311+G(2df,p) hydrogen-bond strength between the dimer consisting of X and N. [b] $\Delta E_{Int5'}$ is defined as the M06-2X/6-31+G(d,p) interaction energy between X and the 5′ flanking base (G for ODN2 or T for ODN1). [c] $\Delta E_{Int3'}$ is defined as the M06-2X/6-31+G(d,p) interaction energy between X and the 3′ flanking base (A for ODN2 or C for ODN1).

bonds compared to Watson-Crick hydrogen bonds. In fact, the (*syn*) adduct base pair is repulsive in ODN1_b:1′(C) (by 6 kJ/mol, Table 6.2), and only weakly stable in ODN2_b:2′(C) (by –24 kJ/mol, Table 6.2), due to the presence of one or a maximum of two hydrogen bonds. Specifically, N4–H of C hydrogen bonds with O6 of the adduct for 48 – 58% of the simulation (Table 6.4). Alternatively, one N4–H of C can form a contact with N7 of the adduct (<6% occupancy, Table 6.4), while the other amino hydrogen is in contact with O6 (15 – 27% occupancy, Table 6.4). The Hoogsteen hydrogen bond involving N7 in this latter conformation competes with the intramolecular O–H···N7 hydrogen bond, which is intact

Table 6.3 Hydrogen-bond Occupancies for X = a over the Duration of MD Simulations

N	Strand	Adduct conformation	H-Bond[a]	%[b]	Avg Distance[c]	Avg Angle[d]
C	ODN1	*anti*	N2–H1···O2	84.5	2.845	16.6
			N1–H···N3	76.0	2.906	15.9
			N4–H1···O6	71.7	2.865	16.8
		syn	N4–H2···O6	53.9	2.846	28.9
			N4–H1···N7	15.4	2.911	44.4
			N4–H1···O6	6.4	2.834	22.8
	ODN2	*anti*	N2–H1···O2	79.2	2.856	15.3
			N1–H···N3	77.0	2.906	14.7
			N4–H1···O6	72.4	2.865	16.3
		syn	N4–H1···O6	31.9	2.840	21.8
			N4–H2···O6	9.0	2.840	35.2
			N4–H1···N7	7.3	2.914	39.4
G	ODN1	*anti*	N2–$H1_a$···N7	49.6	2.902	22.0
			N1–H_a···O6	29.5	2.858	22.9
			N1–H_a···N7	10.4	2.923	28.8
		syn	N1–H···$O6_a$	65.6	2.856	20.7
			N2–H···$N7_a$	54.6	2.899	24.7
	ODN2	*anti*	N1–H_a···O6	56.0	2.844	22.3
			N2–$H1_a$···N7	38.9	2.910	21.9
			N2–$H1_a$···O6	5.6	2.874	32.0
		syn	N1–H···$O6_a$	62.2	2.856	21.1
			N2–H1···$N7_a$	51.3	2.901	23.7
			N2–H1···$O6_a$	5.4	2.873	32.7

[a] Data is provided for hydrogen bonds within 3 Å heavy atom separation and 120° X–H–X angle for greater than 5% of the total simulation time. To distinguish between X:G mismatches, atoms belonging to the adduct are denoted with a subscript "a". [b] Average distance (Å) between heavy atoms when a hydrogen bond is present. [c] Average hydrogen-bond angle (°) as calculated by the AmberTools package, where 0° represents a linear hydrogen bond.

for 70 – 72% of the simulation. Due to the poor hydrogen bonding between the bases, the strands are expected to be destabilized when the *syn* conformation of the adduct is paired with C, compared with strands containing natural dG or the *anti* adduct paired with C. When *p*-PhOH-dG adopts the *syn* conformation opposite C, there are some differences in Hoogsteen bonding compared to *o*-PhOH-dG. Specifically, the Hoogsteen bond between N4–H of C and O6 in the adduct is present for less time than for *p*-PhOH-dG (6% for ODN1 and 31% for ODN2, Table 6.3) compared to *o*-PhOH-dG (48 – 58%, Table 6.4). Instead, the dominant binding conformation of *p*-PhOH-dG in ODN1 involves two hydrogen bonds

Table 6.4 Hydrogen-bond Occupancies for X = b over the Duration of MD Simulations

N	Strand	Adduct conformation	H-Bond[a]	%[b]	Avg Distance[c]	Avg Angle[d]
C	ODN1	*anti*	N1–H···N3	70.8	2.913	15.6
			N2–H···O2	80.3	2.852	17.5
			N4–H1···O6	72.4	2.862	17.4
			3′-N4–H···O11	18.8	2.912	20.4
			OH···N7	6.2	2.809	40.7
		syn	OH···N7	72.6	2.867	42.3
			N4–H1···O6	48.7	2.842	17.2
			N4–H2···O6	27.7	2.844	25.1
	ODN2	*anti*	N2–H···O2	81.7	2.849	16.1
			N1–H···N3	76.0	2.904	16.1
			OH···N7	73.6	2.822	43.5
			N4–H···O6	66.1	2.870	16.5
		syn	OH···N7	70.1	2.867	43.0
			N4–H1···O6	58.5	2.843	20.7
			N4–H2···O6	15.6	2.844	36.9
			N4–H1···N7	6.3	2.913	46.1
G	ODN1	*anti*	N2–$H1_a$···N7	37.0	2.907	22.5
			N1–H_a···O6	28.9	2.844	25.3
			N1–H_a···N7	12.3	2.926	28.6
			N2–$H1_a$···O6	10.2	2.870	32.7
		syn	N1–H···$O6_a$	82.7	2.823	23.6
			OH···N7	74.6	2.856	40.4
			N2–H1···$O6_a$	32.0	2.879	32.8
			N2–H1···$N7_a$	14.6	2.918	39.7
	ODN2	*anti*	N1–$H1_a$···O6	65.2	2.840	23.9
			N2–$H1_a$···N7	36.4	2.901	23.1
			OH···N7	21.6	2.856	44.3
			N2–$H1_a$···O6	16.5	2.856	32.3
			OH···O2P	8.0	2.790	19.8
			OH···O5′	6.4	2.847	32.1
		syn	N1–H···$O6_a$	54.5	2.849	22.3
			N2–H1···$N7_a$	27.9	2.915	28.1
			OH···N7	15.9	2.842	42.6
			N2–H1···$O6_a$	15.6	2.874	32.1
			5′-N2–H2···O11	11.7	2.904	10.99

[a] Hydrogen bonds are considered present if within 3 Å heavy atom separation and 120° X–H–X angle for greater than 5% of the total simulation time. To distinguish between X:G mismatches, atoms belonging to the adduct are denoted with a subscript "a". [b] Average distance (Å) between heavy atoms when a hydrogen bond is present. [c] Average hydrogen-bond angle (°) as calculated by the AmberTools package, where 0° represents a linear hydrogen bond.

between the N4 amino group of C and O6 (53%), as well as interaction between N1–H of C

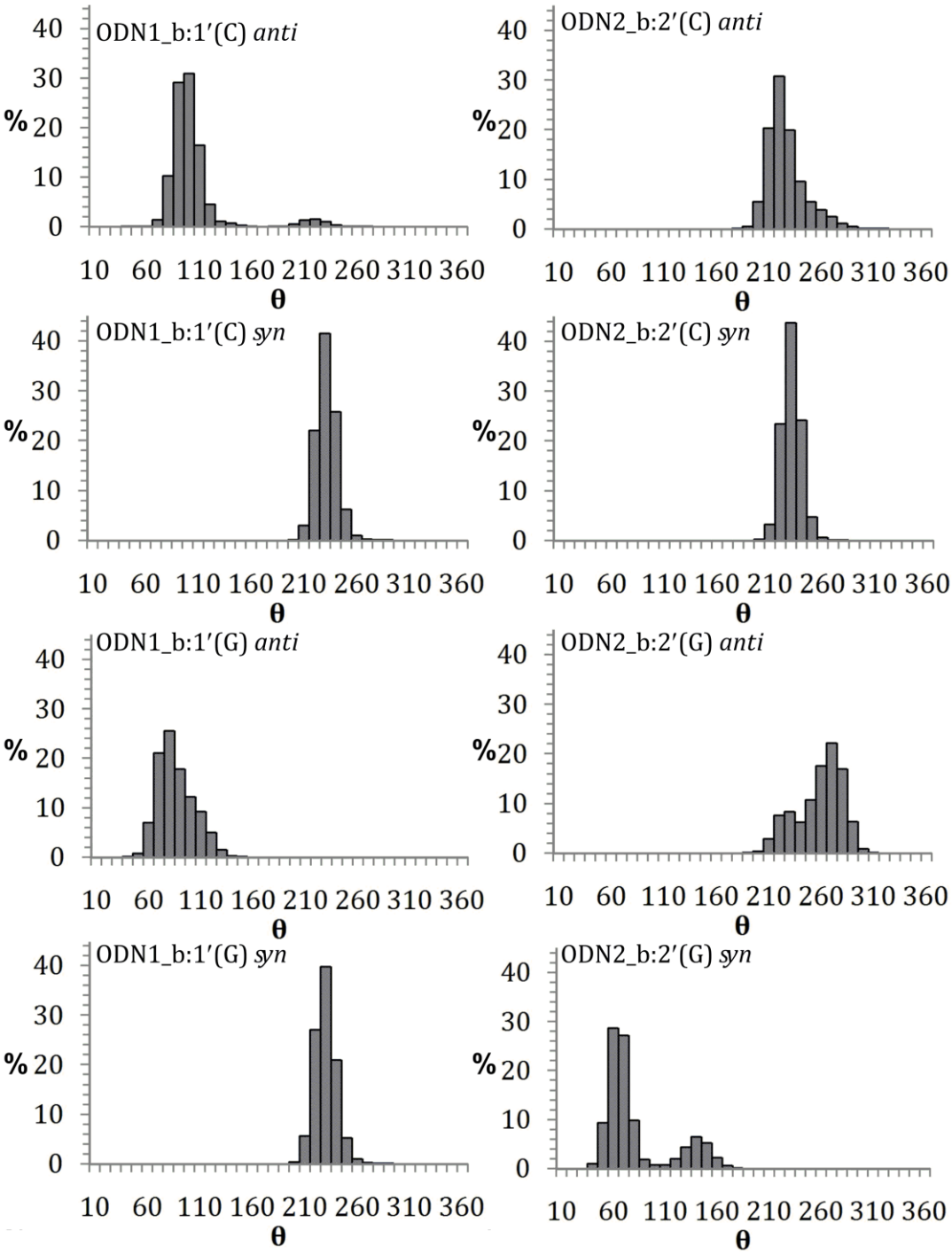


Figure 6.3 Percent distribution of θ (degrees) throughout the 20 ns trajectories for the *o*-PhOH-dG (X = b) adduct.

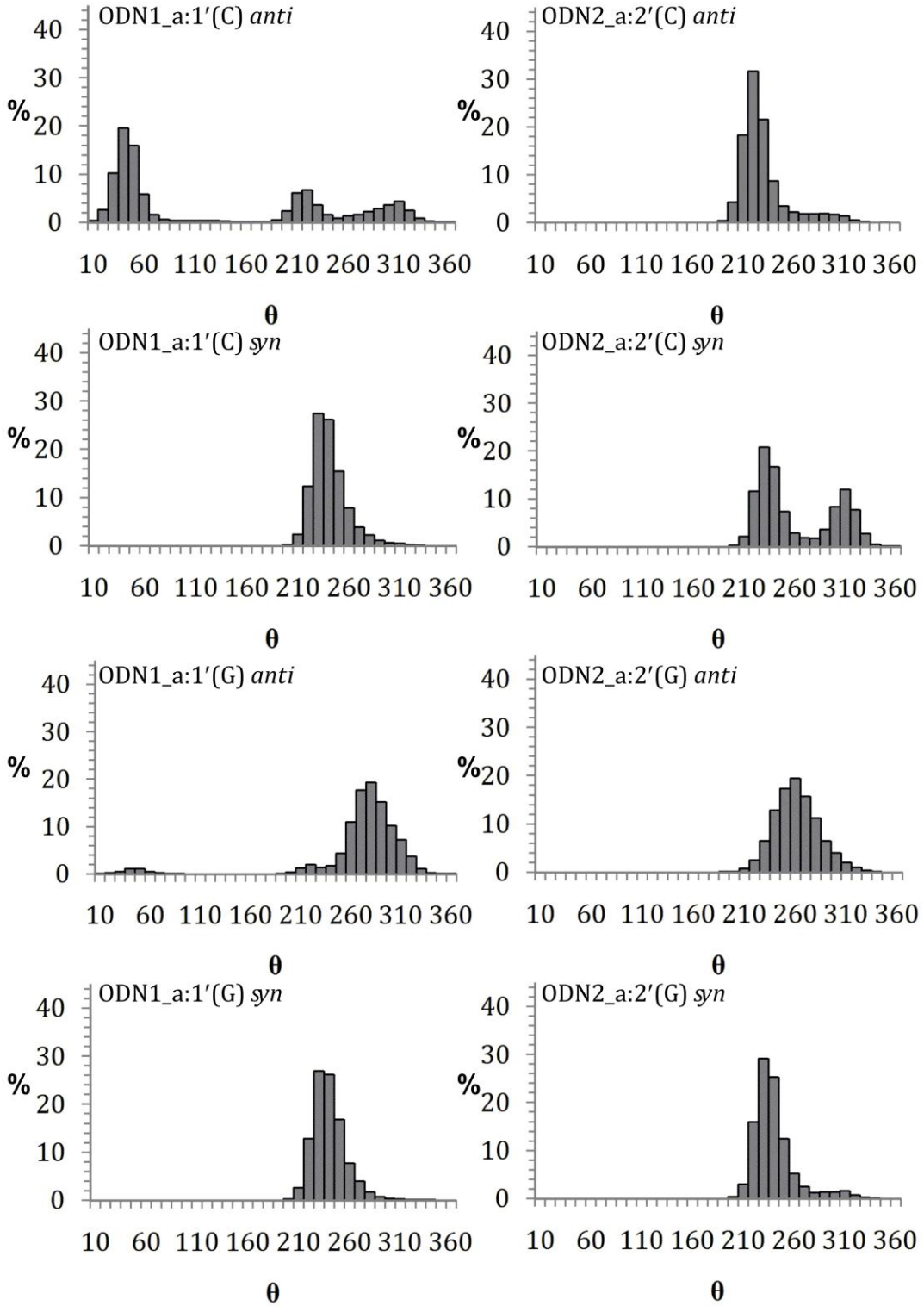


Figure 6.4 Percent distribution of θ (degrees) throughout the 20 ns trajectories for the *p*-PhOH-dG (X = a) adduct.

and N7 (15%). This increased Hoogsteen bonding to N7 in ODN1 is likely due to a lack of competing O–H···N7 hydrogen bond. This results in a base-pair stability of –35 kJ/mol in ODN1 (Table 6.2). The Hoogsteen bond occupancies are less in ODN2 (9% and 7%), which leads to weaker hydrogen bonding (–3.6 kJ/mol, Table 6.2), although the reason for this is currently unclear. The average value of θ twist is 235.7° for ODN1, but is 256.9° for ODN2, which results from two minimum conformations (~230 and 310°, Figure 6.4).

6.3.3 *Anti* Conformation of Adducts (X = a or b) for N = G

When the adducts adopt the *anti* conformation opposite *syn* G, two hydrogen-bond contacts occur between the Watson-Crick face of the adduct and the Hoogsteen face of G. Specifically, the N2 amino group of the adduct bonds to N7 of G (36 – 49%, Tables 6.3 and 6.4), and N1–H of the adduct bonds to O6 of G (28 – 65%). However, the Hoogsteen base pair strength is considerably weaker for the modified ODN1:1′(G) (–23 to –26 kJ/mol compared to –50 kJ/mol, Table 6.2) compared to the unmodified G:G mismatch, while within ODN2:2′(G) the base pairing strength (–49 to –56 kJ/mol, Table 6.2) is comparable to the unmodified G:G mismatch. As found when the *anti* adducts are paired with C, the θ twist is dependent on both the sequence and the adduct. For ODN1_a:1′(G), the average θ is 265.7° with three minimum conformations at θ = ~280, 220, and 50°, and for ODN2_a:2′(G) there is a single conformation (θ = 256.5°, Figure 6.4). This corresponds to a highly twisted structure regardless of sequence for *p*-PhOH-dG. For *o*-PhOH-dG, an average θ value of 81.1° occurs for ODN1_b:1′(G) (Figure 6.3), which corresponds to a highly twisted structure (similar to *p*-PhOH-dG) with no O–H···N7 hydrogen bond (within the cutoff criteria, Table 6.4). For ODN2_b:2′(G), the average θ equals 253.4°, with two minima (Figure 6.3). In the first minimum (θ ~ 230°) the O–H···N7 hydrogen bond is intact (22%, Table 6.4), while the phenoxyl group rotates (θ ~ 270°) later in the simulation to form a bifurcated hydrogen bond with the phosphate group of the same residue (6 – 8%, Table 6.4). These changes in θ

lead to adduct stacking interactions between –13 and –20 kJ/mol with the 5′-flanking base, and –34 and –54 kJ/mol with the 3′-flanking base (Table 6.2). In general, the magnitude of the stacking interactions for the adducts and the unmodified G:G mismatch are similar, although *o*-PhOH-dG shows a greater interaction with the 3′-flanking base within ODN1 compared to the G:G mismatch (–54 versus –43 kJ/mol, Table 6.2). These calculated structures for *anti*-adduct:*syn*-G within ODN1 predict a more stable structure for *o*-PhOH-dG versus *p*-PhOH-dG and that both should be less stable than the unmodified G:G mismatch due primarily to the significant drop in hydrogen bond stability.

6.3.4 *Syn* Conformation of the Adducts (X = a or b) for N = G

When the adducts adopt the *syn* conformation opposite *anti* G, a dramatic drop in Hoogsteen hydrogen-bonding stability within ODN1 does not occur. For *p*-PhOH-dG, two strong (–46 to –48 kJ/mol, Table 6.2) Hoogsteen bonds (51 – 65%, Table 6.3) form between the adduct and G regardless of the sequence, which are only slightly less stable than the unmodified G:G mismatch (–50 kJ/mol, Table 6.2). The duplexes containing *o*-PhOH-dG show a slightly greater decrease in hydrogen-bonding stability (–41, –45 compared to –50 kJ/mol, Table 6.2). For ODN1_b:1′(G), N1–H of the Watson-Crick face of G hydrogen bonds to O6 of the adduct (83%, Table 6.4). In addition, a second N2–H···O6 Hoogsteen bond is sometimes present (32%), while a third Hoogsteen bond (N2–H···N7) is present for less (15%) time. The low percentage of Hoogsteen bonding to N7 may be related to the competing (O–H···N7) hydrogen bond (75%, Table 6.4). For ODN2, the two Hoogsteen bonds to O6 of *o*-PhOH-dG are less occupied (16 – 54%, Table 6.4) than ODN1. However, the N2–H···N7 Hoogsteen bond is more occupied (28%) than ODN1 (15%). This can be correlated to a much higher degree of θ twist than for ODN1 (76.7° compared to 224.6°, Figure 6.3), which corresponds to a structure with the O–H···N7 interaction disrupted and the phenoxyl group forming a N2–H···O interaction (11.7%, Table 6.4) with the 5′-flanking

G base. This reduces the occupation of the O–H···N7 contact (to 16%) and frees N7 to form Hoogsteen bonds.

6.4 Discussion

6.4.1 Structural Properties of C-Linked Phenoxyl Adducts in Duplex DNA: Comparison to Experiment

Experimentally measured melting temperatures of the duplexes studied in this chapter (Table 6.5)[1] indicate that the PhOH-dG adduct destabilizes the duplex when base paired with its normal partner C. This observation is in agreement with previous studies showing that bulky N-linked aromatic amine lesions decrease duplex stability.[20-22] As discussed in Chapter 1, factors thought to cause the destabilizing influence of the N-linked aromatic amine lesions when base paired with C depend on the conformation of the adduct. If the adduct is present in the *anti* conformation, it can form a Watson-Crick base pair with C. This base pairing places the bulky aromatic amine group outside the helix in the major groove where it is not involved in base stacking interactions. Solvent exposure of hydrophobic aryl groups is proposed to destabilize the *anti* conformation.[21,22] If the C8-AA-dG adduct is present in the *syn* conformation, then Watson-Crick base pairing with C is not possible. For example, the N-linked C8-AAF-dG adduct favors a *syn* base-displaced intercalated conformation where the aromatic AAF is intercalated into the helix with local unwinding. The Watson-Crick face of the G moiety in the adduct then projects into the major groove, while the opposing pyrimidine (C) base is flipped out of the helix.[17,18,23] This duplex is considerably distorted compared to the unmodified duplex and is denatured at the lesion site, which typically leads to bulge stabilization and frameshift mutations.

For the phenoxyl C-linked adducts in the *anti* conformation, binding to C should result in three strong Watson-Crick hydrogen bonds (Figure 6.5a) similar to the unmodified sequences, which would cause little destabilization (Table 6.2). The relatively small ring

Table 6.5 T_m Values Derived from UV Melting Experiments

ODN	T_m[a]	ΔT_m[b]
ODN1_a:1′(C)	27	–17
ODN1_a:1′(G)	31	+1
ODN1_b:1′(C)	31	–13
ODN1_b:1′(G)	24	–6
ODN2_a:2′(C)	28	–16
ODN2_a:2′(G)	39	+9
ODN2_b:2′(C)	38	–6
ODN2_b:2′(G)	39	+9

[a] In °C, taken from Reference 19. [b] Change in T_m relative to unmodified duplexes. Taken from Reference 19.

size of these adducts also suggests that solvent exposure would not pose the large energetic penalty observed for aromatic amine or PAH adducts. Thus, the *anti* conformation for phenoxyl adducts is unlikely to account for the significant drop in duplex stability (6 – 17 °C, Table 6.5) when base paired with C. In fact, the decrease in duplex stability for the X:C pair is similar to the decrease noted when a mismatch is introduced into the unmodified duplex. This suggests unfavorable hydrogen-bonding interactions within X:C. Therefore, based on the molecular dynamics study combined with the available melting temperature data,[1] it is predicted that the C-linked phenoxyl adducts adopt the *syn* conformation within both sequences, and the loss of Watson-Crick hydrogen bonds explains the observed decrease in melting temperature (Figure 6.5b).

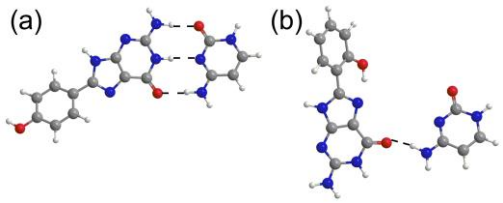


Figure 6.5 Average structures of hydrogen-bonded base pairs from MD simulations: (a) ODN1_a:1′(C) with the *anti* conformation of the adduct; and (b) ODN1_b:1′(C) with the *syn* conformation of the adduct.

For *o*-PhOH-dG, enhanced stacking interactions (Table 6.2) in ODN2_b:2′(C) (*syn* adduct) compared to the unmodified sequence (*anti* G) may explain why ODN2_b:2′(C) is less destabilized than ODN1_b:1′(C), which has weaker stacking relative to the unmodified sequence. Furthermore, the X:C hydrogen bonding is more favorable in ODN2 than in ODN1 (Table 6.2). Comparison of the MD simulations for ODN1_a and ODN2_a to the T_m results leads to similar conclusions noted for strands containing *o*-PhOH-dG. Specifically, the decrease in stability (by 17 and 16 °C, Table 6.5) observed for strands containing the *p*-PhOH-dG:C pair implies that the *syn* conformation of the adduct is present in both duplexes. Unlike *o*-PhOH-dG, there is no improved stability for ODN2_a compared to ODN1_a, which is likely due to cancellation of improved stacking with weakened hydrogen-bonds (Table 6.2). The fact that the adduct prefers to be in the *syn* conformation, despite loss in stability due to weaker hydrogen-bonding, implies that other (sterics, major groove contacts, etc.) factors push the adduct to adopt a *syn* conformation regardless of the base in the opposite strand. This notion is in line with the calculated backbone conformations (Table 6.1), which suggest that the *syn* conformation causes less distortion to the backbone than the *anti* conformation. It is also supported by the *anti*/*syn* relative energies calculated in Chapter 5, which show that the *syn* conformer is preferred regardless of sequence. Experimental evidence regarding the fluorescent properties of the C-linked phenoxyl adducts upon duplex formation[1] also support a *syn* preference, where results suggest a decrease in solvent exposure of the modified base upon duplex formation.[1]

It is interesting that the T_m data show that the G mismatch with *p*-PhOH-dG is more stable than that with *o*-PhOH-dG in ODN1:1′(G). The *syn* *p*-PhOH-dG:G base pair in ODN1:1′(G) involves two hydrogen-bonds between N1–H of G and O6 of the adduct (66%) and N2–H of G and N7 of the adduct (55%) (Figure 6.6a). This is similar to the natural unmodified sequence, where N7 is free to form Hoogsteen bonds. Thus, ODN1_a:1′(G) and

the unmodified G:G mismatch have similar stabilities (ΔT_m = +1°C, Table 6.5). However, the stability of ODN1_b:1′(G) is quite different. Specifically, since O–H···N7 is highly populated (75%), N7 is not as available to form Hoogsteen bonds. In fact, only one Hoogsteen bond (N1–H···O6) is occupied for the majority of the simulation (83%), while N2–H of G bonds to N7 of the adduct only 15% of the time. For some of the simulation (32%), N2–H of G also hydrogen-bonds to O6 (Figure 6.6b). Therefore, *p*-PhOH-dG forms a much more stable base pair with G than *o*-PhOH-dG in ODN1:1′(G). This is not reflected in the calculated hydrogen-bond strengths (Table 6.2) since multiple bonds are occupied in the average structure of ODN1_b:1′(G), even though this is not the case for the majority of the simulation (Figure 6.6c). The same arguments also explain why ODN1_b:1′(G) is destabilized with respect to the natural G:G mismatch (by –6°C, Table 6.5).

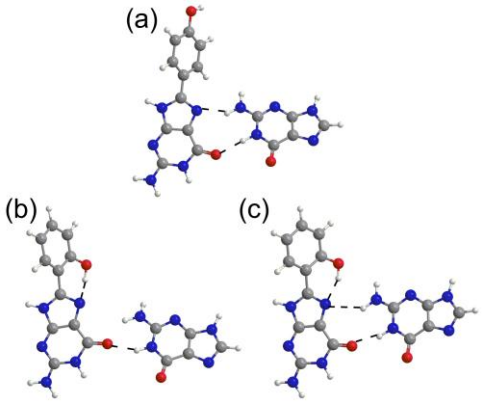


Figure 6.6 Hydrogen-bonding in ODN1:1′(G). (a) The *syn* conformation of *p*-PhOH-dG paired with G. (b) A representative structure of the hydrogen-bonding pattern for the *syn* conformation of *o*-PhOH-dG paired with G. (c) The average hydrogen-bonding pattern of the *syn* conformation of *o*-PhOH-dG paired with G that overestimates the strand stability.

The stability of *p*-PhOH-dG:G and *o*-PhOH-dG:G base pairs is sequence dependent. ODN2_a:2′(G) and ODN2_b:2′(G) are stabilized to the same degree (+9°C) with respect to the unmodified G:G mismatched sequences. Close examination of the base pairs throughout

the simulation reveals that two Hoogsteen N1–H···O6 (62%) and N2–H···N7 (51%) *p*-PhOH-dG hydrogen-bonds are not affected by the sequence (Figure 6.7a). However, *o*-PhOH-dG forms two reasonably strong Hoogsteen bonds in ODN2_b:2′ (Figure 6.7b) rather than just one in ODN1_b:1′ (Figure 6.6b). Specifically, N1–H···O6 (54%) and N2–H···N7 (28%) are both present in ODN2_b:2′, which is more similar to the bonding in *p*-PhOH-dG and the unmodified base (35 – 66%). In ODN1_b:1′(G), N7 is blocked from forming stable bonds due to the stability of O–H···N7 (75%). However, in ODN2_b:2′(G), the O–H···N7 bond is present only 16% of the simulation, which allows for Hoogsteen bonding to occur during the remainder of the simulation. This effect is sequence specific since the hydrogen bond is broken due to formation of a weak N2–H···O contact (12%) between the 5′-flanking G and the phenoxyl oxygen (Figure 6.7c), which distorts θ (76.7°). This provides a possible explanation for the observed relative stabilities of ODN2_a:2′(G) and ODN2_b:2′(G) (Table 6.2).

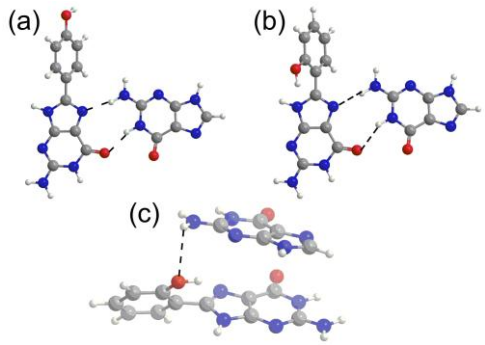


Figure 6.7 Hydrogen bonding in ODN2:2′(G). (a) The *syn* conformation of *p*-PhOH-dG paired with G. (b) The *syn* conformation of *o*-PhOH-dG paired with G. (c) The hydrogen-bonding interaction between *o*-PhOH-dG and the 5′-flanking G base.

Furthermore, the calculations provide insight into experimental observations regarding the lack of tautomerization in duplexes compared to the isolated nucleoside adduct (Figure 2.2).[1] Keto tautomers for duplexes involving ODN2_b are not detected,[1] even though *o*-PhOH-dG is predicted to be in the *syn* conformation and is not solvent exposed

(see Chapter 2, Figure 2.2). This is likely due to the interaction of the phenoxyl OH in *o*-PhOH-dG with the 5′-flanking G in ODN2 observed in MD simulations. Studies on structurally related systems that undergo ESIPT have reported similar observations, where the inability to detect the keto tautomer in a 2-(2′-hydroxyphenyl)benzoxazole (HBO) artificial nucleoside[24] was attributed to formation of a stable hydrogen bond between the phenoxyl OH and the 3′-flanking sugar O4′ atom.[24]

The increase in stability of the modified G:G mismatch within ODN2 may be due to the enhanced stacking of the *syn* adduct with purine bases compared with the pyrimidine bases (Table 6.2). Stronger stacking arises since the phenoxyl group is involved in intrastrand interactions with the flanking bases in ODN2 (Figures 6.8 a and b), but to a lesser extent in ODN1. Indeed, the pyrimidine bases are too small to overlap with the entire adduct, and therefore only interact with dG in a manner similar to the natural strand (Figures 6.8 c and d). The stacking strengths calculated for the *syn* conformations of the adducts (Table 6.2) support enhanced stability of ODN2:2′ upon incorporation of both adducts. This is also supported by experimental evidence for greater fluorescence quenching of the C-linked phenoxyl adducts within ODN2,[1] suggesting more favorable stacking interactions with the flanking purine bases.

6.4.2 Comparison to Previous Small Model Results

It is interesting to compare the results from the MD simulations on a duplex model to the DFT results using various small models discussed in the previous chapters and thereby evaluate the ability of the small models to predict structures relevant to DNA. First, the structure of the adduct base pairs can be directly compared to the small model hydrogen-bonding study presented in Chapter 3. As discussed in Section 6.1, G and C were selected as the complementary bases in the strands considered in the MD simulations since the small model results imply these pairs result in the most stable structures. Indeed, this is

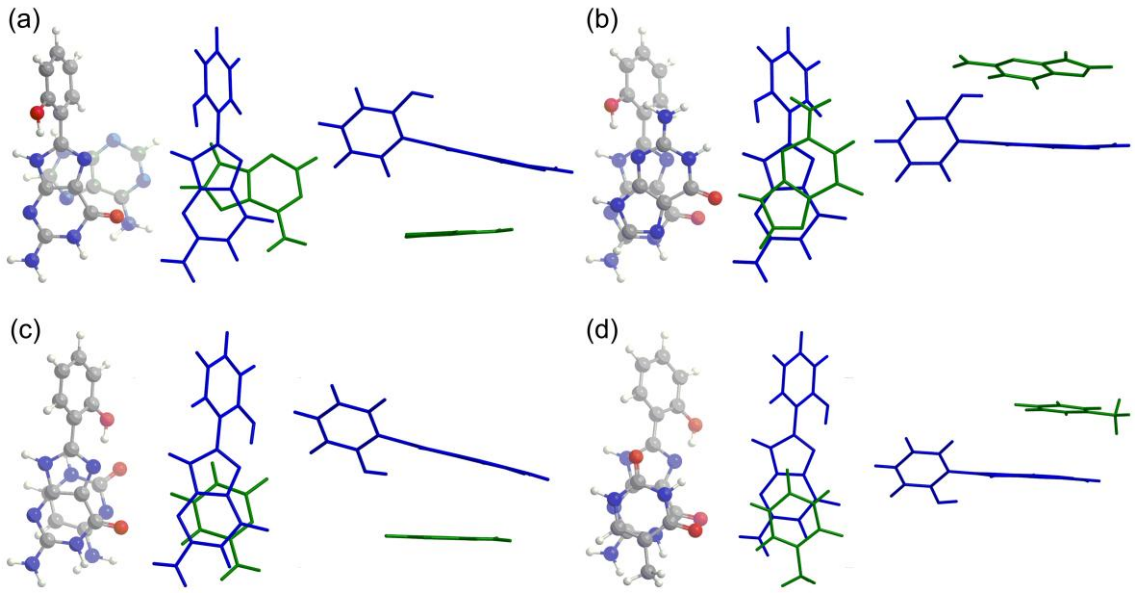


Figure 6.8 Face (left, center) and edge (right) views of the intrastrand stacking interactions between *o*-PhOH-dG and the 3′- (a, c) or 5′- (b, d) flanking base in the ODN2:2′ (a, b) and ODN1:1′ (c, d) sequences.

confirmed by experimental melting temperatures.[1] The DFT-optimized structures in Chapter 3 have interaction strengths for C Hoogsteen bonded to *o*-PhOH-dG or *p*-PhOH-dG of –98 kJ mol^{-1} and –54 kJ mol^{-1}, respectively, while the corresponding interaction strengths for the MD structures are 6 – 35.5 kJ mol^{-1}. It is apparent that the binding strengths of the adducts to C are greatly overestimated by the small model. This arises because the hydroxyl group of *o*-PhOH-dG remains hydrogen-bonded to N7 for the majority of the simulation (Figure 6.5) rather than interacting with C, despite similar values of θ for MD and DFT structures. This blocks N7 from participating in Hoogsteen bonds with C in the MD structures, and thus only one of the four bonds predicted by the DFT calculations (Figures 3.4 and 3.6) forms in the MD simulation. This difference arises due to the effects of the helix environment on the base pair arrangement. Specifically, while structures were free to re-orient to maximize hydrogen-bonding interactions during the DFT-optimizations, steric constraints prevent the same interactions in the MD simulations. Indeed, the

hydrogen bonds in the MD simulations are so weak that it is not surprising DFT leads to quite different optimized structures.

For the adduct:G Hoogsteen pair, the interaction strengths of the DFT-optimized structures are found to be –57 and –53 kJ mol^{-1} for the *o*- and *p*-PhOH-dG adducts, respectively (Chapter 3), which is only 5 – 16 kJ mol^{-1} more stable than predicted by the MD structures (–41 to – 48 kJ mol^{-1}). The hydrogen-bonding arrangement predicted by DFT is also remarkably similar to the MD structures of ODN2. In particular, the MD structures have two strong Hoogsteen bonds, which is the most stable bonding arrangement in the DFT optimizations as well (Figures 3.4 and 3.6 compared to Figure 6.7). This arises since the orientation of the hydroxyl group in the *o*-PhOH-dG adduct calculated by DFT, where the O–H···N7 hydrogen bond is disrupted, matches the MD results. However, the MD structures have much less buckle and propeller distortion than the DFT structures, due to differences in the surrounding environment, where the MD structures are flanked by 5′ and 3′ base pairs. The planarity of the MD structure compared to the DFT structure may explain the slight decrease in calculated stability.

Overall, the DFT hydrogen bonding model provides accurate estimates of the most stable base pairs possible in MD simulations. However, when steric effects in the helix prevent the most stable base pairs from forming in MD simulations, DFT-optimized structures overestimate the stability of hydrogen-bonded base pairs. Therefore, the structural criteria used in Chapter 3 to predict biologically-relevant base pairs is insufficient when the base pairs are so unstable that they undergo significant distortion to maximize hydrogen-bond contacts.

In addition to preferred bonding patterns, it is of interest to determine whether small models can accurately predict the preferred conformation and structural features of bulky adducts in DNA helices. Comparisons will be made between the deoxydinucleoside

monophosphate model in Chapter 5 and the MD results, since this small model was the largest used to represent the DNA environment, and no significant changes occurred in the adduct structure compared to smaller models. Some differences between the deoxydinucleoside monophosphate and MD structures arise in the base–base orientation, as well as values of θ, due to the presence of both the 5′ and 3′ flanking bases and the backbone. Additionally, there is less ring-ring overlap with flanking bases in MD simulations (Figure 6.8) than predicted by the deoxydinucleoside monophosphate model (Figure 5.11), which suggests the isolated deoxydinucleoside monophosphate model overestimates the importance of these interactions. However, the hydroxyl group can form discrete interactions with the flanking bases in MD structures as predicted by the smaller model. In fact, the interactions are stronger in the MD structures due to an increased twist, as well as the base opposite the adduct which competes with the OH group for the N7 hydrogen-bond site. Regardless of the structural differences, it is significant that the deoxydinucleoside monophosphate model is able to successfully predict the *anti*/*syn* preference of the adducts in DNA, which implies that the role of steric interactions with the flanking residue are crucial for predicting the conformational stability. Therefore, this is the smallest model that should be used when studying the *anti*/*syn* conformational preferences of bulky C8 adducts.

6.4.3 Biological Implications of C-linked Phenoxyl Adducts

As discussed in Chapter 1, the mutagenicity of C8 adducts is strongly related to the conformational preferences.[23] For the *syn* conformation of C8-AA-dG adducts, and the *anti* conformation of N2-PAH-dG adducts, the bulky group can be either stacked in the helix or wedged in the minor groove. The bulky group in *syn* C-linked phenoxyl adducts most closely resembles the stacked conformation adopted by N-linked adducts, where the phenoxyl group participates in stacking with the flanking bases. However, the relatively

small size of the phenoxyl group compared to AA or PAH aromatic ring systems results in some conformational differences. The MD simulations suggest that the C8-phenyl group is well-accommodated opposite C within the helix and the C base is not flipped out, as noted for AA and PAH adducts.[23] This observation is important in terms of biological activity, as it suggests that the C-linked phenoxyl adducts will be less able to stabilize bulges that induce frameshift mutations. The MD simulations also suggest that the C-linked phenoxyl adducts minimally perturb the B-form duplex DNA structure regardless of the conformational preference of the adduct within the helix. As discussed in Chapter 5, these findings suggest that the C-linked phenoxyl adducts will be less repair-prone than other adducts,[23] since they do not cause major distortion to the duplex in the *syn* base-displaced intercalated conformer.

While the biological activity of C-linked phenoxyl adducts has yet to be addressed, Chapter 1 reports results of primer extension assays for C8-Ph-dG within a single 24-mer oligonucleotide with the lesion flanked by pyrimidine bases (C).[25] UV thermal melting experiments were also performed for C8-Ph-dG within the 12-mer sequence 5′-d(GCGCCXGCGGTG), where X = C8-Ph-dG.[25] Unfortunately, the sequence used for the thermal melting data did not match the sequence used for the primer-extension assays.[25] Within the 12-mer, the T_m of the duplex revealed the ability of 8-Ph-dG to stabilize a G:G mismatch, as reported for the phenoxyl adducts. Depending on the polymerase used, correct incorporation or blocks to replication could occur. In addition, small amounts of G and A nucleotides were incorporated opposite the lesion and two-base deletions were also observed, suggesting that C8-Ph-dG is weakly mutagenic and capable of generating G→C and G→T transversions and deletions in cells.[25] The present chapter reveals evidence for G→C transversion mutations to help explain the potential carcinogenicity of C8-PhOH-dG adducts.

6.5 Conclusion

The MD results presented in this chapter demonstrate the structural impact of C-linked phenoxyl adducts within two decanucleotides. Experimental evidence shows that the C8-dG adducts decrease duplex stability (6 – 17°C)[1] when base paired with C, but show a sequence dependent increase in duplex stability when mismatched with G (–6 – 9°C).[1] MD structures provide important insights into these experimental results. Specifically, experiments alone could not determine the conformation adopted by the adducts in DNA, but melting temperature observations combined with the changes in stability for the *anti* and *syn* conformers predicted using structures obtained from MD simulations support a *syn* preference of the adducts in both duplexes. This demonstrates the usefulness of combining calculations with experiments to understand the structural and biological impact of DNA lesions. MD and DFT calculations also predict stabilization of the G:G mismatch by the C-linked phenoxyl adducts, where the stabilizing effect is expected to be greater in the decanucleotide containing flanking purine bases due to stronger base–base interactions. These results are further confirmed by experimental melting temperatures.

The *syn* preference of the adducts, coupled with their ability to stabilize G:G mismatches in a sequence specific fashion suggests the possibility for C-linked phenoxyl adducts to display mutagenic hotspots, as previously noted for the corresponding N-linked C8-dG adducts derived from arylamine carcinogens.[26] Specifically, since the phenoxyl adduct lesions are better able to stabilize G:G mismatches when flanked by purine bases, purine-rich sites would be expected to show a greater propensity for misincorporation of G opposite the adduct. Future experiments should test this prediction by incorporating the C-linked phenoxyl adducts into the purine-rich *Nar*I oligonucleotide. Similar to studies of C8-Ph-dG,[25] primer-extension assays should be used to determine the biological impact of the adducts in this sequence. This will also allow for comparison to the N-linked aromatic

amine adducts, which have been well-studied within this sequence context.[21,27-31] However, the work in this chapter indicates that experiments alone will not be enough to fully understand the conformational and base-pairing preferences of the adducts in *Nar*I oligonucleotides. MD simulations will be crucial in order to explain and aid interpretation of the experimental results.

6.6 References

(1) Omumi, A.;[§] Millen, A.L.;[§] Wetmore, S.D.; Manderville, R.A. *Chem. Res. Toxicol.* **2011**, submitted for publication. ([§] These authors contributed equally to this work).

(2) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817-3829.

(3) Heavner, S.; Gannett, P. M. *J. Biomol. Struct. Dyn.* **2005**, *23*, 203-219.

(4) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graph.* **2006**, *25*, 247-260.

(5) Dupradeau, F. Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821-7839.

(6) Pigache, A.; Cieplak, P.; Dupradeau, F. Y. *Abstr. Pap. Am. Chem. Soc.* **2004**, *227*, 181-COMP.

(7) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845-862.

(8) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.;

Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03, Revisions C.02 and D.01; Gaussian, Inc.: Wallingford, CT, 2004.

(9) D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, P. A. Kollman. AMBER 10: University of California, San Francisco, 2008.

(10) D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. P. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. A. Kollman. AMBER 11: University of California, San Francisco, CA, 2010.

(11) GaussView, Version 5, Roy Dennington, Todd Keith and John Millam, Semichem Inc., Shawnee Mission KS, 2009.

(12) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.

(13) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577-8593.

(14) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327-341.

(15) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215-241.

(16) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.;

Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Jr., J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; Revision A.02 ed.; Gaussian, Inc.: Wallingford CT, 2009.

(17) Patnaik, S.; Cho, B. P. *Chem. Res. Toxicol.* **2010**, *23*, 1650-1652.

(18) Ohandley, S. F.; Sanford, D. G.; Xu, R.; Lester, C. C.; Hingerty, B. E.; Broyde, S.; Krugh, T. R. *Biochemistry* **1993**, *32*, 2481-2497.

(19) Schneider, B.; Neidle, S.; Berman, H. M. *Biopolymers* **1997**, *42*, 113-124.

(20) Valis, L.; Mayer-Enthart, E.; Wagenknecht, H. A. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 3184-3187.

(21) Elmquist, C. E.; Wang, F.; Stover, J. S.; Stone, M. P.; Rizzo, C. J. *Chem. Res. Toxicol.* **2007**, *20*, 445-454.

(22) Koehl, P.; Valladier, P.; Lefevre, J. F.; Fuchs, R. P. P. *Nucleic Acids Res.* **1989**, *17*, 9531-9541.

(23) Patel, D. J.; Mao, B.; Gu, Z. T.; Hingerty, B. E.; Gorin, A.; Basu, A. K.; Broyde, S. *Chem. Res. Toxicol.* **1998**, *11*, 391-407.

(24) Dupradeau, F. Y.; Case, D. A.; Yu, C. Z.; Jimenez, R.; Romesberg, F. E. *J. Am. Chem. Soc.* **2005**, *127*, 15612-15617.

(25) Kohda, K.; Tsunomoto, H.; Kasamatsu, T.; Sawamura, F.; Terashima, I.; Shibutani, S. *Chem. Res. Toxicol.* **1997**, *10*, 1351-1358.

(26) Roy, D.; Hingerty, B. E.; Shapiro, R.; Broyde, S. *Chem. Res. Toxicol.* **1998**, *11*, 1301-1311.

(27) Belguisevalladier, P.; Fuchs, R. P. P. *J. Mol. Biol.* **1995**, *249*, 903-913.

(28) Mao, B.; Hingerty, B. E.; Broyde, S.; Patel, D. J. *Biochemistry* **1998**, *37*, 95-106.

(29) Wu, X. Y.; Shapiro, R.; Broyde, S. *Chem. Res. Toxicol.* **1999**, *12*, 895-905.

(30) Fuchs, R. P. P.; Koffel-Schwartz, N.; Pelet, S.; Janel-Bintz, R.; Napolitano, R.; Becherel, O. J.; Broschard, T. H.; Burnouf, D. Y.; Wagner, J. *Biochem. Soc. Trans.* **2001**, *29*, 191-195.

(31) Wang, F.; Elmquist, C. E.; Stover, J. S.; Rizzo, C. J.; Stone, M. P. *Biochemistry* **2007**, *46*, 8498-8516.

7 Chapter 7: Conclusion

7.1 Thesis Overview

This thesis systematically developed a computational model to predict the conformational and base-pairing preferences of the C-linked PhOH-dG adducts by initially considering the nucleobase and nucleoside adducts and gradually incorporating the backbone, flanking bases, and complementary base. This approach provided insight into the effects of each factor on the conformational preferences of the adducts and thus added value to the experimental results available for some of these models. Finally, all of the structural features were combined in a single model to study the structure and stability of DNA duplexes containing phenoxyl damaged lesions.

The initial work on small (nucleobase, nucleoside) models was used to complement ongoing experimental investigations of the structure and stability of the phenoxyl adducts. The calculations not only matched the experimentally-predicted structure of the adduct in a variety of solvents, but explained observations regarding differences in stability of the glycosidic bond for the *o*- and *p*- adducts. In particular, a twisted *syn* conformation of the nucleoside structure was predicted by calculations, which matched experimental observations. However, calculations revealed that the reason for the *syn* preference is the presence of an O5′–H···N3 hydrogen bond. This implied that the experimental observations of the *syn* preference of the nucleoside adduct cannot necessarily be extrapolated to a biological system. Furthermore, the adducts are predicted to be stable in DNA, which makes study of their conformational properties crucial to understanding their mutagenicity.

The potential for mispairing due to the conformational flexibility of the adducts was investigated in Chapter 3 to provide insight into the possible base-pairing preferences of the adducts in the absence of experimental data. While the *anti* conformation exhibited a preference for base-pairing with cytosine, the *syn* conformation formed stable base pairs

with both cytosine and guanine. Due to the possibility of different Watson-Crick and Hoogsteen base-pairing preferences, which could lead to mutations depending on the conformation adopted in DNA, it was important to determine the conformational preferences of the adducts using more biologically-relevant models than described in Chapter 2 and used in early experiments. Therefore, Chapter 4 expanded the model to include the phosphate component of the backbone. Ultimately, this model predicted a greater stability for the *anti* conformation of the adducts, where small calculated energy differences suggested the possibility of a complex conformational heterogeneity in DNA helices, as found for other C8-adducts.[1-9]

Since the conformational flexibility of other adducts has been found to be sequence dependent,[5,10-15] and steric clash is expected to be an important factor in determining the stability of the *anti* conformation of bulky adducts in DNA,[16] the model was further expanded in Chapter 5 to include the 5′- or 3′-flanking base. This led to a reversal in the relative energies of the *anti* and *syn* conformations compared to the nucleotide model. Indeed, the *syn* conformation is preferred with this model for both adducts regardless of the identity of the flanking base. This suggested that the neighboring sugar and backbone play an important role in destabilizing the *anti* conformation of phenoxyl adducts and that the *syn* conformation may be more relevant to DNA. Therefore, nucleotide models are insufficient to study *anti*/*syn* conformational preferences of bulky adducts, and the *syn* conformation is important to consider in predicting possible mismatches in DNA.

In Chapter 6, MD simulations of both the *anti* and *syn* conformations of the adducts paired with cytosine or guanine in two DNA sequences were performed to provide greater insight into the conformational and base-pairing preferences of the adducts in biologically-relevant systems. The calculations were able to explain experimental observations regarding duplex stability, sequence dependence, and base-pairing preferences.

Furthermore, the simulations predicted that the adducts preferentially adopt a *syn* conformation in DNA regardless of sequence or the identity of the base in the complementary strand. This agreed with results obtained using a deoxydinucleoside monophosphate model, and therefore this is the smallest QM model that should be used to study the *anti*/*syn* conformational preferences of bulky C8 adducts. However, nucleotide and nucleoside models were able to provide insight into other questions, such as the preferred orientation of the bulky group with respect to dG, sugar puckering, and preferred values of χ for both the *anti* and *syn* conformations. Therefore, many insights into adduct structure and base-pairing preferences can be readily obtained by calculations using small nucleoside, nucleotide, and deoxydinucleoside monophosphate models in the absence of experimental data.

Furthermore, information regarding the biological implications have been obtained. Little distortion of the backbone was observed in the simulations, which suggested that repair enzymes may overlook the phenoxyl adduct lesions. The results also predicted that Hoogsteen base-pairing leads to significant destabilization of the helix when the adduct is opposite C, and therefore mismatches are possible for the adducts. In particular, mismatch stabilization was predicted to occur when the adduct is Hoogsteen bonded to guanine, which rationalized measured relative duplex stabilities. The evidence for mismatch stabilization indicates the possibility that guanine may be preferentially inserted opposite the C-linked phenoxyl adducts, which would lead to G→C mutations upon replication, as observed for structurally-related (Ph-dG) adducts. Therefore, the results of this thesis have implications for the general carcinogenic model of C8 adduct genotoxicity and help develop a broad picture rationalizing the observation of G→C mutations for similar (aryl hydrazine)[17] DNA lesions.

Overall, this computational approach provided deeper insight into the results obtained by experimental methods, and successfully guided the development of new experiments. For example, calculations suggested duplexes containing C and G in the complementary strand opposite the adduct site should be synthesized and melting temperatures determined due to the superior binding strengths of these pairs (Chapter 3). Furthermore, the interpretation of the computational results was guided by experimental data. For example, the differences in measured melting temperatures combined with the structural data obtained from MD simulations revealed the preferred conformation of the adducts. Therefore, calculations are a valuable tool to guide and complement experimental studies of mutagenic lesions in DNA.

7.2 Further Applications of the Methodology Applied in this Thesis

The methodology applied in this thesis can be readily extended to other C8-modified nucleobases. Indeed, work using a combination of the approaches in this thesis to study structurally-related systems has already been published.[18] Specifically, C8-guanine derivatives (pyrrole and indole-linked guanine) have been synthesized using a nucleoside model due to their potential utility as fluorescent probes of Watson-Crick versus Hoogsteen hydrogen bonding. Association reactions between the C8-pyrrole and C8-indole-linked guanine compounds with G and dC showed these compounds to be useful fluorescent reporters of hydrogen-bonding specificity. DFT calculations on the structures of the modified bases and the Watson-Crick and Hoogsteen hydrogen-bond strengths between the *syn* or *anti* conformations of the pyrrole- and indole-linked bases were employed to determine the potential base-pairing preferences. Insight into the structural features of the modified bases were obtained from DFT calculations, as done for the phenoxyl adducts in this thesis. Like the phenoxyl adducts, *syn* structures are more stable than *anti*, and contain an O5'–H···N3 hydrogen bond. Calculations provided information about both

the structure and strength of the hydrogen-bonding interactions compared to the natural G:G and G:C base pairs. Both probes were shown to form stable Watson-Crick base pairs with dC and three-point reverse Hoogsteen base pairs with G. The hydrogen-bond strengths of the modified bases paired with guanine are less than the natural Watson-Crick G:C pair, but greater than the unmodified G:G mismatch due to the extra hydrogen-bond contact with the Hoogsteen face of the modified base. Thus, the approach to conformational searching and modeling the hydrogen bond strengths undertaken in Chapters 2 and 3 has been successfully applied to predict experimentally observed bonding arrangements for other C8-modified guanine nucleosides.

Experiments on other C8-modified bases in DNA duplexes have also been performed. Specifically, benzothiophene-linked guanine bases were inserted into DNA helices and their fluorescence properties studied. Calculations using the conformational searching methodology described in Chapter 2, the hydrogen-bonding study in Chapter 3, and the molecular dynamics simulations of the synthesized duplexes (Chapter 6) are in progress to complement these experimental results, where the computational model in each case mirrors the corresponding experimental model. Together, the computational and experimental approaches have provided a greater understanding of a variety of C8-modified systems, and should be used in the future to explore structurally-related systems, where the model used is dependent on the available experimental data and the questions to be answered.

7.3 Future Work

While the work in this thesis has answered many questions regarding the structure and stability of phenoxyl adducts, much remains to be done. In particular, the importance of intrastrand interactions in predicting duplex stability suggests that they should be studied even more thoroughly in the future. Therefore, the significance of flanking bases in

determining adduct conformation, base-pairing properties, and duplex stability should be investigated in a variety of additional sequences. The *Nar*I sequence is particularly interesting due to the wealth of previous literature that has used this sequence to study adduct conformation.[10,14,19-22] This provides ample opportunities to draw comparisons to previous literature to obtain a global understanding of the structural dependence of adducts on sequence context. Therefore, experiments should be performed to synthesise and measure the stability of these modified strands. In addition, molecular dynamics simulations should be performed to provide greater insight into the structural features of the modified *Nar*I sequences and aid interpretation of the experimental results.

In the future, the *in vivo* biological effects of these adducts should be addressed. Therefore, the possibility of guanine mismatch stabilization must be further evaluated in environments more relevant to cells. Specifically, while G:G mismatches have been found to stabilize synthetic duplex DNA, the stability of the mismatch in a replication fork environment remains to be determined. It will be essential to identify the preferred base pairing in this environment, as has been done for other adducts,[23,24] since it will be the most relevant for determining the identity of the nucleobase incorporated in the complementary strand upon replication. Furthermore, it will be especially important to understand how these adducts interact with polymerase enzyme active sites. Different polymerases have different fidelities and are known to result in different mutations for similar adducts. Therefore, experiments should be performed to determine whether polymerases stall in the presence of the adducts, whether translesion bypass synthesis is undertaken, and whether translesion synthesis is error-prone. These experiments will guide the development of new computational models, where molecular dynamics simulations have successfully modeled various polymerase active sites with bulky adducts to determine which nucleobase is preferentially incorporated opposite the lesion.[4,25-27] These future directions for study will

ultimately confirm whether the mechanism for G→C transversion mutations proposed in this work is correct, or whether new models for the genotoxicity of phenoxyl adducts need to be developed.

Future work should include studies of the conformational and binding preferences of other substituted phenyl adducts, such as the aryl hydrazine adducts discussed in Chapter 1 which have only been studied in Z-DNA, using the methodology described in this thesis. Specifically, a nucleoside model can reveal important information regarding the degree of twist in the nucleobase, preferred sugar puckering, and the presence of strong intramolecular interactions. A hydrogen-bonding base pair study using nucleoside models will identify the base pairs to be considered in larger models. A deoxydinucleoside monophosphate model will reveal important intrastrand interactions for consideration in larger models. MD simulations of the adducts in DNA can provide valuable information regarding the *anti*/*syn* conformational and base pairing preferences of the adducts in DNA. This study of aryl hydrazine adducts will reveal whether the identity of the phenyl substituent affects the mutagenicity of C8-phenyl adducts, and begin to develop a comprehensive model for carbon-bonded C8-aryl-dG adduct genotoxicity.

7.4 Global Conclusions

DNA damage is important to understand since it may lead to cancer if unrepaired. In particular, bulky C8 guanine adducts are known to induce a variety of mutations due to the conformational flexibility of the adducts in DNA. C-linked C8-PhOH-dG adducts are an important type of DNA damage that is poorly understood. This thesis developed a computational model for the study of the conformational and base-pairing preferences of phenoxyl adducts, and validated the model through comparison to experimental results so it can now be applied to other bulky C8-modified guanine derivatives. The structure of the adducts in DNA have been determined, where the bulky group induces a *syn* conformation

in DNA duplexes similar to other C8 adducts. A stabilized guanine mismatch that is also sequence-dependent has been identified for the *syn* adducts, which suggests that the primary mechanism of genotoxicity may be base-substitution mutations resulting in G→C transversions. The small size of the phenoxyl adducts compared to other bulky C8 adducts leads to small structural changes in the helix, which supports the hypothesis that these lesions may be poorly recognized by DNA repair mechanisms. This thesis has contributed to a growing body of computational and experimental literature dedicated to understanding the role of conformational heterogeneity in determining the mutagenicity of bulky C8 adduct lesions.

7.5 References

(1) Cho, B. P.; Beland, F. A.; Marques, M. M. *Biochemistry* **1994**, *33*, 1373-1384.

(2) Cho, B. P.; Zhou, L. *Biochemistry* **1999**, *38*, 7572-7583.

(3) Cho, B. S. P. *J. Environ. Sci. Health Pt. C-Environ. Carcinog. Ecotoxicol. Rev.* **2004**, *22*, 57-90.

(4) Zhang, L.; Shapiro, R.; Broyde, S. *Chem. Res. Toxicol.* **2005**, *18*, 1347-1363.

(5) Cai, Y.; Patel, D. J.; Geacintov, N. E.; Broyde, S. *J. Mol. Biol.* **2007**, *374*, 292-305.

(6) Meneni, S.; Shell, S. M.; Zou, Y.; Cho, B. P. *Chem. Res. Toxicol.* **2007**, *20*, 6-10.

(7) Liang, F. T.; Cho, B. P. *Biochemistry* **2010**, *49*, 259-266.

(8) Patnaik, S.; Cho, B. P. *Chem. Res. Toxicol.* **2010**, *23*, 1650-1652.

(9) Liang, F. T.; Cho, B. P. *Chem. Res. Toxicol.* **2011**, *24*, 597-605.

(10) Belguisevalladier, P.; Fuchs, R. P. P. *J. Mol. Biol.* **1995**, *249*, 903-913.

(11) Shibutani, S.; Fernandes, A.; Suzuki, N.; Zhou, L.; Johnson, F.; Grollman, A. P. *J. Biol. Chem.* **1999**, *274*, 27433-27438.

(12) Huang, X. W.; Colgate, K. C.; Kolbanovskiy, A.; Amin, S.; Geacintov, N. E. *Chem. Res. Toxicol.* **2002**, *15*, 438-444.

(13) Yang, T. L.; Huang, Y. H.; Cho, B. P. *Chem. Res. Toxicol.* **2006**, *19*, 242-254.

(14) Wang, F.; Elmquist, C. E.; Stover, J. S.; Rizzo, C. J.; Stone, M. P. *Biochemistry* **2007**, *46*, 8498-8516.

(15) Jain, N.; Meneni, S.; Jain, V.; Cho, B. P. *Nucleic Acids Res.* **2009**, *37*, 1628-1637.

(16) Vongsutilers, V.; Phillips, D. J.; Train, B. C.; McKelvey, G. R.; Thomsen, N. M.; Shaughnessy, K. H.; Lewis, J. P.; Gannett, P. M. *Biophys. Chem.* **2011**, *154*, 41-48.

(17) Kohda, K.; Tsunomoto, H.; Kasamatsu, T.; Sawamura, F.; Terashima, I.; Shibutani, S. *Chem. Res. Toxicol.* **1997**, *10*, 1351-1358.

(18) Schlitt, K. M.; Millen, A. L.; Wetmore, S. D.; Manderville, R. A. *Org. Biomol. Chem.* **2011**, *9*, 1565-1571.

(19) Mao, B.; Hingerty, B. E.; Broyde, S.; Patel, D. J. *Biochemistry* **1998**, *37*, 95-106.

(20) Wu, X. Y.; Shapiro, R.; Broyde, S. *Chem. Res. Toxicol.* **1999**, *12*, 895-905.

(21) Fuchs, R. P. P.; Koffel-Schwartz, N.; Pelet, S.; Janel-Bintz, R.; Napolitano, R.; Becherel, O. J.; Broschard, T. H.; Burnouf, D. Y.; Wagner, J. *Biochem. Soc. Trans.* **2001**, *29*, 191-195.

(22) Elmquist, C. E.; Wang, F.; Stover, J. S.; Stone, M. P.; Rizzo, C. J. *Chem. Res. Toxicol.* **2007**, *20*, 445-454.

(23) Gu, Z. T.; Gorin, A.; Hingerty, B. E.; Broyde, S.; Patel, D. J. *Biochemistry* **1999**, *38*, 10855-10870.

(24) Perlow, R. A.; Broyde, S. *J. Mol. Biol.* **2001**, *309*, 519-536.

(25) Perlow, R. A.; Broyde, S. *J. Mol. Biol.* **2002**, *322*, 291-309.

(26) Wang, L. H.; Broyde, S. *Nucleic Acids Res.* **2006**, *34*, 785-795.

(27) Benitez, B. A. S.; Arora, K.; Balistreri, L.; Schlick, T. *J. Mol. Biol.* **2008**, *384*, 1086-1097.

A. Appendix A

Table A.1 Values of backbone torsional angles (degrees) in deoxydinucleoside monophosphates containing natural dG[a]

	Cytosine		Thymine		Guanine		Adenine		Average	
	5′	3′	5′	3′	5′	3′	5′	3′	5′	3′
	anti									
χ5′	242.1	227.8	244.8	218.9	253.5	253.5	251.9	256.2	248.1	239.1
χ3′	237.8	287.1	218.1	280.1	265.3	265.3	270.8	266.8	248.0	274.8
γ5′	48.1	50.9	49.8	50.6	49.7	49.7	46.6	49.8	48.6	50.2
δ5′	146.4	144.9	148.2	137.2	149.0	149.0	146.8	149.7	147.6	145.2
ε5′	189.6	202.9	186.4	199.5	188.4	188.4	195.7	187.6	190.0	194.6
ζ	275.7	283.5	276.9	277.8	277.0	277.0	277.8	277.7	276.9	279.0
α	295.7	291.6	295.7	297.5	297.1	297.1	290.5	296.6	294.8	295.7
β3′	165.9	173.6	164.1	174.3	166.2	166.2	171.1	166.8	166.8	170.2
γ3′	55.4	51.7	56.6	58.6	53.1	53.1	46.7	51.6	52.9	53.7
δ3′	112.3	148.0	86.6	152.8	137.1	137.1	137.1	138.3	118.3	144.0
ε3′	188.1	176.4	193.4	176.2	180.5	180.5	180.9	180.5	185.7	178.4
	syn									
χ5′	76.2	224.0	88.0	225.3	88.5	239.7	86.7	236.8	84.9	231.5
χ3′	239.8	78.1	253.3	78.8	263.9	66.5	261.4	65.2	254.6	72.2
γ5′	45.7	48.3	48.8	48.1	49.2	48.4	49.8	48.5	48.4	48.3
δ5′	138.2	138.3	139.3	140.5	142.9	141.9	143.1	142.2	140.9	140.7
ε5′	165.0	201.1	158.5	199.1	159.4	179.8	149.0	182.4	158.0	190.6
ζ	265.2	282.9	265.9	282.3	266.3	268.6	268.7	272.5	266.5	276.6
α	298.1	294.9	303.1	295.2	301.2	306.9	304.4	306.5	301.7	300.9
β3′	185.9	173.0	180.2	172.3	179.6	171.0	185.5	167.8	182.8	171.0
γ3′	60.4	50.9	62.5	51.2	65.5	55.0	60.0	55.9	62.1	53.3
δ3′	143.7	143.6	146.9	143.4	152.8	136.5	151.4	133.8	148.7	139.3
ε3′	176.5	176.4	176.5	176.4	174.2	179.1	175.6	179.8	175.7	177.9

[a] M06-2X/6-31G(d,p) optimized geometries. All angles are defined in Figure 5.1.

Table A.2 Values of backbone torsional angles (degrees) in deoxydinucleoside monophosphates containing the *o*-PhOH-dG adduct[a]

	Cytosine		Thymine		Guanine		Adenine		Average	
	5′	3′	5′	3′	5′	3′	5′	3′	5′	3′
	anti									
χ5′	241.6	222.4	238.0	236.2	254.3	260.6	245.3	262.7	244.8	245.5
χ3′	252.0	257.2	244.1	256.7	256.5	245.8	267.5	248.6	255.0	252.1
γ5′	54.1	51.5	55.1	51.2	55.0	49.4	52.4	52.4	54.1	51.1
δ5′	87.9	145.8	87.4	147.4	87.5	150.6	81.0	152.5	85.95	149.1
ε5′	166.2	215.0	154.5	207.6	162.3	191.1	170.8	191.1	163.4	201.2
ζ	274.7	290.6	275.3	281.6	273.6	275.1	266.0	276.5	272.4	280.9
α	311.8	276.1	318.6	284.3	315.2	293.8	310.5	293.6	314.0	287.0
β3′	168.5	159.6	167.3	159.8	164.7	160.6	169.7	159.5	167.6	159.9
γ3′	56.9	52.7	61.0	53.9	58.1	57.5	60.9	57.8	59.2	55.5
δ3′	139.9	92.1	142.1	91.5	145.9	90.7	152.6	91.6	145.1	91.5
ε3′	178.8	194.9	178.0	193.5	176.7	193.7	174.5	193.7	177.0	193.9
	syn									
χ5′	84.0	225.1	94.3	220.9	97.3	225.9	68.9	228.3	86.1	225.1
χ3′	250.4	89.8	241.1	89.9	271.1	90.9	279.5	92.8	260.5	90.9
γ5′	47.5	48.6	48.5	47.9	50.0	48.0	45.9	48.3	48.0	48.2
δ5′	139.1	140.7	140.3	140.9	145.0	143.7	134.3	144.2	139.7	142.4
ε5′	174.4	199.6	169.8	199.5	165.9	199.8	179.2	198.6	172.3	199.4
ζ	270.9	284.2	266.4	283.7	263.8	283.5	265.2	284.5	266.6	284.0
α	301.1	292.2	303.2	292.4	303.1	291.8	308.2	290.6	303.9	291.8
β3′	168.9	172.9	171.5	171.5	178.3	172.4	164.1	173.6	170.7	172.6
γ3′	65.5	50.1	68.2	50.8	64.4	50.1	67.1	50.2	66.3	50.3
δ3′	138.3	145.8	140.6	145.3	153.9	146.3	153.9	146.9	146.7	146.1
ε3′	177.5	175.8	177.4	175.8	175.6	176.0	175.1	175.7	176.4	175.8

[a] M06-2X/6-31G(d,p) optimized geometries. All angles are defined in Figure 5.1.

Table A.3 Values of backbone torsional angles (degrees) in deoxydinucleoside monophosphates containing the *p*-PhOH-dG adduct.[a]

	Cytosine		Thymine		Guanine		Adenine		Average	
	5′	3′	5′	3′	5′	3′	5′	3′	5′	3′
	anti									
χ5′	246.1	227.3	241.5	234.0	247.0	232.9	248.1	240.1	245.7	233.6
χ3′	247.8	261.2	245.7	264.0	261.1	263.3	248.0	280.7	250.7	267.3
γ5′	55.5	50.6	54.8	52.5	55.2	51.5	45.2	51.7	52.7	51.6
δ5′	86.1	144.0	86.6	150.0	85.5	149.8	137.6	150.6	99.0	148.6
ε5′	160.4	219.8	153.4	207.7	167.4	214.6	187.0	199.9	167.1	210.5
ζ	276.2	289.3	275.5	287.3	273.0	295.2	225.2	284.6	262.5	289.1
α	313.1	279.0	316.6	283.7	315.2	275.6	304.5	290.1	312.4	282.1
β3′	166.7	158.0	167.1	163.2	165.1	163.2	177.6	168.1	169.1	163.1
γ3′	57.8	51.6	61.2	52.4	61.5	51.7	53.0	52.6	58.4	52.1
δ3′	136.4	90.0	142.1	89.5	151.3	89.7	151.3	140.8	145.3	102.5
ε3′	179.3	193.3	177.7	194.6	175.2	193.9	176.0	178.5	177.1	190.1
	syn									
χ5′	91.4	224.1	93.9	220.3	89.6	228.5	88.3	232.4	90.8	226.3
χ3′	236.2	82.2	242.7	86.5	264.3	87.4	251.2	95.5	248.6	87.9
γ5′	47.8	49.6	48.4	48.5	48.5	48.0	47.5	50.1	48.1	49.1
δ5′	142.2	142.7	139.8	140.8	140.7	143.9	141.0	147.9	140.9	143.8
ε5′	170.9	195.6	169.1	199.2	170.5	198.8	169.4	204.1	170.0	199.4
ζ	257.5	285.0	257.0	283.0	258.8	283.7	250.4	293.7	255.9	286.4
α	300.7	295.3	305.3	292.6	305.0	292.2	305.7	284.2	304.2	291.1
β3′	175.1	173.9	171.5	171.3	170.6	173.1	172.2	169.9	172.4	172.1
γ3′	61.6	49.7	63.7	51.4	65.7	50.4	60.8	54.4	63.0	51.5
δ3′	140.7	144.2	142.9	144.1	151.5	145.1	148.2	146.4	145.8	145.0
ε3′	176.5	175.0	176.1	176.2	175.9	175.8	176.1	175.3	176.2	175.6

[a] M06-2X/6-31G(d,p) optimized geometries. All angles are defined in Figure 5.1.